U.PORTO

FEP FACULDADE DE ECONOMIA UNIVERSIDADE DO PORTO

Hayekian Reflections on Economic Explanation

Diogo de Melo Lourenço

Ph.D. in Economics

Supervisor:

Mário da Graça Moura

2015

Biographical Note

I was born on July 12, 1988 in Porto, Portugal. In 2010 I concluded my undergraduate studies in Economics at the Faculty of Economics of the University Of Porto (FEP). I immediately enrolled in the PhD program in Economics at the same faculty.

At present I lecture Research Methodology to undergraduate students in Communication Sciences at the Faculty of Arts of the University of Porto (FLUP), and Economic History to undergraduate students in Economics at FEP.

Acknowledgments

My deepest gratitude goes to my supervisor, Mário da Graça Moura, whose unabated enthusiasm, incisive criticism and all gracious advice shall never be forgotten. For helping me grow as a thinker and as a person, I thank you, professor.

I am grateful to Pedro Nuno Teixeira, for his friendship and indispensable reality-checks, to Helena Santos, for her warmth and affection, and to Manuel Luís Costa, for his support and creative insights.

To my other friends and family, who have endured the blows of my frustration with fortitude, and the negligence of my absence with patience, all I can offer is my sincere affection. Among these I wish to single out Maria Francisca Peixoto, Maria José Cruz, Tânia Porto Ferreira, Filipe Vicente, Pedro Varela and Leonel Silva.

I thank my friends and colleagues at FEP, namely Nuno Gonçalves, Mariana Cunha, Luís Guimarães, Filipe Grilo, Carlos Seixas, Duarte Leite and Vera Rocha. I believe we are veritable Battle Buddies.

I also owe a debt of gratitude to Luís Terrinha and João Abreu for their presence during important stretches of my graduate work.

Finally, financial support by the Foundation for Science and Technology (FCT) in the form of a PhD grant (SFRH / BD / 71701 / 2010) made this work possible.

Resumo

Esta tese é uma reflecção Hayekiana sobre questões fundamentais para a ciência económica. No capítulo 2 apresento observações psicológicas e filosóficas que considero necessárias para o entendimento económico da agência. Os capítulos 3 e 4 incidem sobre a obra *Scientism and the Study of Society* de Hayek, aludindo também a outras obras do autor publicadas no mesmo período. No capítulo 3 estudo a sua conceção de conhecimento nas ciências naturais e concluo que a sua tentativa de distinguir ciência de senso-comum leva-o a uma distinção entre esquema e conteúdo difícil de sustentar. Dedico o capítulo 4 ao pensamento de Hayek sobre as ciências morais. Clarifico o conceito de ação e avalio a tese de Hayek de que as ciências morais não a explicam. Hayek considera que o cientista moral necessita entender os agentes e que, para isso, deve munir-se da sua experiência enquanto ser humano. A minha tentativa de elucidar esta posição leva-me a concluir a sua ambiguidade. Por fim, discuto duas características que Hayek considera distintivas do trabalho nas ciências morais: a natureza funcional dos seus predicados, e a lógica das frases que envolvem atitudes proposicionais (por ex. acreditar que, saber que, etc.). Concluindo que Hayek não é construtivo no que toca aos procedimentos a seguir para entender os agentes, inicio o capítulo quinto notando que o problema do cientista moral é o problema de todos nós: como entender os outros com uma base evidencial limitada e comportamental. Associo estas questões às motivações históricas das teorias da preferência revelada e da decisão. Neste capítulo, importo uma lição do domínio da pragmática, enfatizo a indeterminação das interpretações intersubjetivas, e aponto as implicações normativas dessa indeterminação. O último capítulo procura trazer uma nova luz ao Debate sobre o Cálculo no Socialismo aplicando as lições dos capítulos anteriores. Aqui, articulo três proposições que creio estarem frequentemente identificadas nas contribuições mais relevantes para o debate: a universalidade da teoria económica, a determinação objetiva dos fundamentais, e a visão desses fundamentais enquanto existindo fora de qualquer contexto social. Termino com algumas reflexões preliminares sobre empreendedorismo e coordenação intersubjetiva num contexto descentralizado.

Abstract

The work that follows is a reflection, from a Hayekian standpoint, on fundamental questions of economics. Chapter 2 presents philosophical and psychological theses that I believe to be necessary for the economic understanding of agency. Chapters 3 and 4 both focus on Hayek's *Scientism and the Study of Society,* complemented by other works of Hayek's from the same period. I begin in chapter 3 with his conception of inquiry in the natural sciences. I find in Hayek an attempt to differentiate science from common-sense that leads him to an untenable scheme-content distinction. In chapter 4, I turn to his views on the moral sciences. I clarify the notion of action, and I study Hayek's claims that the moral sciences do not *explain* action, concluding that his position requires qualification. Hayek argues that the moral scientist needs to understand people and that, for this, she must tap her experience as a thinking human. I try to figure out what this claim involves and show that it does not amount to a definite position. I then discuss two features Hayek seems to have found distinctive about moral scientific explanation, viz. the functional nature of its predicates and the logical form of sentences involving the propositional attitudes (e.g. to believe *that*, to know *that*, etc.). After arguing in chapter 4 that Hayek does not explain how the moral scientist is to understand agents, I begin chapter 5 by noticing that the moral scientist's predicament is everyone's predicament: how to *understand* other agents based on a limited, behavioural evidential basis. I relate these matters to the historical motivation of important work in economics, viz. of revealed preference and decision theories. I study these two fields on the lookout for insights on how agents develop theories of each other. I draw a lesson from the subject of pragmatics, emphasize the indeterminacy of intersubjective interpretation, and notice the normative implications of such indeterminacy. The final chapter utilizes the lessons of the previous two to cast a new light on the Socialist Calculation Debate. I articulate three propositions that I believe to underlie the work of the most relevant contributors to the debate: the universality of economic *theory,* the objectivity of economic fundamentals, and a view of these fundamentals as existing outside of any social context. I bring the thesis to a close by offering a few tentative reflections on entrepreneurship and intersubjective coordination in a decentralized setup.

Contents

1. Introduction

I do utterly disapprove and declare against that pernicious custom of making the preface a bill of fare to the book. For I have always looked upon it as a high point of indiscretion in monstermongers and other retailers of strange sights to hang out a fair large picture over the door, drawn after the life, with a most eloquent description underneath. This has saved me many a threepence, for my curiosity was fully satisfied, and I never offered to go in, though often invited by the urging and attending orator with his last moving and standing piece of rhetoric, "Sir, upon my word, we are just going to begin." Such is exactly the fate at this time of Prefaces, Epistles, Advertisements, Introductions, Prolegomenas, Apparatuses, To the Readers's.

Jonathan Swift in A Tale of a Tub, section V

If he were alive today, Swift would keep on disapproving: such custom did not die in the eighteenth century, nor did it limit itself to literature. When referring to academic writing, Deirdre McCloskey (1999, p. 37) calls such custom "boilerplate" and pleads: "Don't, please, please, for God's sake, don't. Nine out of ten readers skip to the substance, if they can find it."

There is something to be said for their disapproval. A Bill of Fare is often oppressive, in that it makes authorial intention inordinately imposing. A great writer does not need it: everything is clear, transparent, and well-articulated. But, alas, one has not been endowed with Swift's talent, or McCloskey's rhetorical powers. One has to make up for what one lacks. So the reader will forgive me if I indulge in a practice whose excuse is the sanction of custom.

The work that lies ahead is a methodological reflection on economic explanation from a Hayekian viewpoint. My reflections started when I wrote a piece on the Socialist Calculation Debate of the 1930s during my first years as a graduate student. This is, of course, the episode that made Austrian economists aware of their differences from other

schools issued from the Marginalist Revolution. Indeed, as late as 1932, Ludwig von Mises still saw those differences as mostly terminological (cf. Kirzner (1988, p. 9)). The study of this debate led me to a close reading of Friedrich von Hayek's work, especially that of the 1930s, 40s and early 50s. I found myself digesting not only classics like his *Economics and Knowledge* (1937), *Scientism and the Study of Society* (1942-4), and *The Use of Knowledge in Society* (1945), but also his *The Sensory Order* (1952), and *The Pure Theory of Capital* (1941).

There was something fascinating about Hayek. I found his subjectivism *very* persuasive. His first-person view of economic reality suited me. It is a position that, it still seems to me, offers a worldview that dignifies the person: it emphasizes the subject's autonomy and the constitutivity of rationality to action. It entreats you to try hard to *understand* other people, even when they do something that seems foolish.

Hayek's subjectivism teaches us that coordination is a matter of people's plans of action dovetailing, and that these plans depend on their beliefs and worldviews: for economics, the way agents see the world is at least as important as the factual correctness of their beliefs. Moreover, Hayek persuasively argues for the autonomy of the economic. Indeed, whether a wrench is a means of production or a box of chocolates a consumption good is not reducible to the structural properties of those entities, but depends on the *role* they have in someone's plans. It is through decentralized, competitive social processes that people's plans and decisions are coordinated.

Persuaded by Hayek, I began doubting the coherence of certain notions, such as that of *data* somehow determinate outside of any plan, or of *objective* conditions of production. But these notions left a void in their stead: apparently there was nothing *there* to coordinate. What if plans, decisions and worldviews do not exist prior to the processes that supposedly coordinate them, but are themselves created within those very processes? Could not coordination and creation be two faces of the same social processes of intersubjective interaction? But then what *are* these processes? *How* do people, with unique causal histories, subjective worldviews, and situated contexts of action coordinate? How do they develop expectations of each other's behaviour? How does Paris get fed? Why are people ever *right?* As Hayek (1937, p. 34) wrote, "before we can explain why people commit mistakes, we must first explain why they should

ever be right." Without fundamentals, it was as though the invisible hand was, as it were, running on empty. I found myself going from solipsism to scepticism. I was also not sure of the ontological implications of these positions.

Meanwhile, I had been writing a piece on what I called, after David Ruccio and Jack Amariglio (2003), 'postmodern moments' in economics. After being introduced to McCloskey's work by reading Mäki (1995) as an undergraduate, this article left me with the impression that there was something about that whole line of thought (which I then, on good authority, called 'postmodernism') that I did not understand. Mäki spoke a lot about truth and justification, and McCloskey spoke a lot about Richard Rorty. It thus made sense to investigate theories of truth and to read Rorty ([1979] 2009).

There were two names that kept catching my attention: Donald Davidson and Willard Van Quine. I had heard *of* Quine, but knew little about him. Davidson was completely unknown to me. But he should not have been. He had published in *Econometrica*, and, together with Patrick Suppes, is one of the pioneers of experimental decision theory (cf. Davidson & Suppes (1956) and Davidson & Suppes (1957)) . Both Quine and Davidson, from what I could gather at the time, seemed to discuss issues that were related to what had left me adrift with Hayek. These two authors discussed how people form worldviews and how people, in particular contexts, integrate what they observe fellow agents do, and listen them say, into expectations about agents' future behaviour - which seemed of a piece with Hayek's emphasis on the coordination of subjective expectations. I also learned that they discussed the notion of rationality, inquiry, ontology and the relation between the mental and the physical. *Each* a topic of fundamental interest to the Hayekian scholar.

I thus ventured into the den of contemporary philosophy. The view was not inviting to a young, perplexed, economist: it is a highly technical, formal, and specialized subject. But, so I hope, this philosophical education has allowed me to throw light upon the economic matters that I found puzzling.

The work that follows is a reflection, from a Hayekian standpoint, on fundamental questions of economics: the nature of explanation in the moral sciences, the relation between common-sense and (moral)scientific-sense, what action is and how it can be

explained, the costs and benefits of individualism, the particularities of the logic of economic notions and operators, what exists in the economic domain of reality, and how (entrepreneurial) creativity is possible.

Chapter 2 introduces and presents philosophical and psychological theses that I believe to be necessary for the economic understanding of agency. It draws on the work of Willard Quine, Donald Davidson, Nelson Goodman, Richard Rorty and the later Wittgenstein. I start with a discussion of the variety of available worldversions, showing that there are many *interesting* ways of talking about the world. I then study how individuals, with nothing more than the irritation of their sensory surfaces, adopt and develop worldversions. My reflections point to the importance of contextualizing explanation, and I make this idea explicit by appealing to Wittgenstein's notion of language games. This allows me to better show not only that there are many interesting worldversions, but also that they often belong to different (social) contexts, and thus need not be intersubstitutable.

I end this first chapter with a section on ontology. My emphasis on the variety of possible worldversions may have left a smack of relativism. In trying to clarify the implications of this position, I engage in ontological reflections head on. I offer grounds for rejecting a view of truth as correspondence to a ready-made world, but I defend the non-epistemic features of our intuitions about truth. This echoes Tony Lawson's (1997) position that there is a transitive and an intransitive dimension to knowledge. I also argue that the various vocabularies and theories are somewhat like tools in a toolbox: they should be judiciously chosen depending on the questions we wish to see answered, and on our goals.

These philosophical points settled, I am in a position to engage with Hayek's work of the 1940s. My goal is twofold: to present an unapologetic Hayekian view of economics, whilst at the same time *clarifying* Hayek's own positions. I try to strike a balance between going beyond and going into his views.

Chapters 3 and 4 both focus on his *Scientism and the Study of Society*, complemented whenever necessary by other works of Hayek's from the same period. They are, therefore, part of a long line of work studying this work of Hayek's (cf. Caldwell (2005,

Appendix D)). I begin in chapter 3 with his views on inquiry in the natural sciences. These are critically analysed in detail because they prove crucial to understanding his views on the moral sciences. I find in Hayek an attempt to differentiate science from common-sense that leads him, albeit implicitly, to an untenable scheme-content distinction.

In chapter 4, I turn to his views on the moral sciences. I clarify the notion of action, and present a Davidsonian and broadly common-sense view of reasons as causes of action, which I believe to represent some of Hayek's intuitions well. I study Hayek's claims that the moral sciences do not *explain* action, concluding that his position requires qualification. Hayek argues that the moral scientist needs to understand people, and that, in order to do this, she must tap her experience as a thinking human. I try to figure out what his claim involves that the moral scientist shares a mental structure with her subjects, showing that the claim does not amount to a definite position.

I then discuss two features Hayek seems to have (implicitly) found distinctive about moral scientific explanation, viz. the functional nature of its predicates and the logical form of sentences involving the propositional attitudes (e.g. to believe *that*, to know *that*, etc.). I notice that neither is distinctive of the moral sciences, but that the specificity of the mental vocabulary the moral scientist employs *might* be. I end this chapter with a discussion of Hayek's individualism. I argue that his strain of individualism is sound, not implying implausible ontological or methodological theses.

After arguing in chapter 4 that Hayek does not explain in detail how the moral scientist is to understand agents, I begin chapter 5 by noticing that the moral scientist's predicament is everyone's predicament: how to *understand* other agents based on a limited, behavioural evidential basis. Thus understood, I can relate these matters to the historical motivation of important work in economics, viz. of revealed preference and decision theories. Focusing on the work of Paul Samuelson, Hendrik Houthakker, and Marcel Richter, but also of Frank Ramsey, and Leonard Savage, I study these two fields on the lookout for insights on how agents develop theories of each other. My reflections are mostly critical: I believe that it is in the failure of these approaches that the most valuable lessons lie. One of the criticisms I wage is that these theories attempt to understand action in a contextual vacuum, by way of closed axioms, ever more

sophisticated. I draw a lesson from the subject of pragmatics, emphasize the indeterminacy of intersubjective interpretations, and notice its normative implications.

The final chapter utilizes the lessons of the previous two to cast a new light on the Socialist Calculation Debate. I articulate three propositions that I believe to underlie the work of the most relevant contributors to the debate: the universality of economic *theory*, the objectivity of economic fundamentals, and a view of these fundamentals as existing outside of any social context. I offer an interpretation of the Misesian challenge to socialism, I study the answer of Oskar Lange, and Lerner's contribution to the debate, and I look into Mises and Hayek's work on the division of knowledge for elucidation on matters of fundamental economic theory. I bring the thesis to a close by offering a few tentative reflections on entrepreneurship and intersubjective coordination in a decentralized setup that follow from the insights of the previous two chapters, viz. the autonomy of the economic, and the openness of economic reality.

2. Metatheoretical Preliminaries

I believe that the constellation of Orion is real. If you do not believe me, just look up and see! *There* it is[1]! However, I have been told that I am wrong, that constellations do not exist, that what I call constellations are *nothing but* celestial bodies, stars, planets and what-not, so many light-years away from each other. That we see constellations is simply due to their shapes being salient to the eye, which connects the dots, and paints the night sky with images of Gods, animals, and episodes rife with human significance. Yet, we could wonder whether notions such as that of planet, light-year, or eye are not themselves a result of human creativity, if they are anything else than humans' connecting dots in particular ways, structuring them in a certain fashion; delineating entities or attributing properties in ways that they find significant.

When Nelson Goodman titled his (1978) book *Ways of Worldmaking*, he implicated that there is, or could be, more than one way of worldmaking. Indeed, Goodman urges that there is "a vast variety of versions and visions in the several sciences, in the works of different painters and writers, and in our perceptions as informed by these, by circumstances, and by our own insights, interests, and past experiences." (p. 3)[2]

Like Goodman, I think it difficult to doubt that there are many world-versions, many ways of connecting dots. In fact, there is no need to leave the economics department to find striking examples. There are economists, Hayek for instance, who looked at the market economy as a non-teleological process of intersubjective coordination, whilst others, e.g. the market socialists I will discuss in chapter 6, saw it as a mechanism tending towards an end, a resting point, an equilibrium. Where some have seen harmony of interests (e.g. A. Smith ([1790] 1984, Part IV Ch. 1, p. 184)[3]), others have seen the

[1] If you are suitably located, and there are no clouds and light pollution, of course. But the constellation is there whether you see it or not.

[2] There is thus, conceivably, much idiosyncrasy in worldviews, resulting as they do from our histories and sensitivities. With our worldmaking, we make sense of ourselves, our associates and our surroundings. Our worldviews are expressed through and in every action and interaction in which we engage. The notion of "vision" is also central in Schumpeter's conception of the scientific process, where it is connected with "ideology". See Schumpeter ([1954] 1994, Ch. 4).

[3] "in spite of their [of the rich] natural selfishness and rapacity, though they mean only their own conveniency, though the sole end which they propose from the labours of all the thousands whom they

contradictions of capitalist society (e.g. Marx ([1885] 2006, Vol II Ch. 20, p. 487)[4]). Where some have thought it desirable to let social processes run their course (e.g. V. L. Smith (2007, p. 324)[5]) others have asked for policy to fine-tune the economy (e.g. Robinson (1962, p. 138)[6]). Needless to say, that there are many versions does not imply that there are many *interesting* versions. Yet, I think that Goodman is right when he says that "many different world-versions are of independent interest and importance." (p. 4)

Not only are many versions of independent interest, but agents' idiosyncratic worldmaking is crucial to the understanding of economic action and calculation. As Hayek (1937, p. 37) writes, "the assumptions or hypotheses, which we have to introduce when we want to explain the social processes, concern the relation of the thought of an individual to the outside world, the question to what extent and how his knowledge corresponds to the external facts." Accordingly, it is necessary to reflect on the etiology of worldversions, i.e., on how *human* beings develop and adopt world-versions, i.e. on how they *worldmake*. The goal of this chapter is precisely to supply preliminary reflections on these issues to which I can appeal in the following chapters.

In 2.1. I explore the way the human mind learns, viz. by discriminating and associating stimuli. In 2.2. I discuss *how* such discrimination and association are done by addressing the riddles of induction. I get out of the individual mind by emphasizing the importance of the linguistic practices in an agent's community to the understanding of her inductive tendencies. My perspective is naturalistic in the sense that I will take it as axiomatic that human beings' sole source of evidence about their surroundings comes

employ, be the gratification of their own vain and insatiable desires, they divide with the poor the produce of all their improvements."

[4] "It thus appears that capitalist production involves certain conditions independent of people's good or bad intentions [Willen], which permit the relative prosperity of the working class only temporarily, and moreover always as a harbinger [Sturmvogel] of crisis."

[5] "More generally, in the larger society the rules themselves also emerge as institutions in a spontaneous order – they are found, not deliberately designed by only one calculating mind. Initially constructivist institutions undergo evolutionary change adapting beyond the circumstances that gave them birth. What emerges is a form of "social mind" that solves complex organization problems without conscious cognition. This social mind is born of the interaction among all individuals through the rules of institutions that have to date survived cultural selection processes."

[6] "If more total saving is needed than would be forthcoming under *laisser faire* it can easily be supplemented by budget surpluses."

from the stimulation of their sensory surfaces[7]. As Quine (1973, p. 2) notices, our problem "is that of finding ways, in keeping with natural science, whereby the human animal can have projected this same science from the sensory information that could reach him according to this science."[8] In 2.3. I get even further out of the individual mind, and draw on Wittgenstein's late work to discuss the importance of context in making sense of the vast diversity of vocabularies and theories that we employ every day. In 2.4. I diffuse some unpalatable ontological implications that might seem to follow from the positions I adopt in the prior sections.

2.1. The Mind and Worldmaking

In *The Sensory Order*, Hayek (1952) adopted a somewhat similar perspective to the naturalism I described. He tried to give a materialistic understanding of the mind: to elucidate how the order that is the mind could, in principle, arise from the intercourse of the nervous system with its surroundings. In his (1952, p. 47) own words, "how is it possible to construct from the known elements of the neural system a structure which would be capable of performing such discrimination in its responses to stimuli as we know our mind in fact to perform."

An intuitive first answer might be to say that differences between stimuli cause similarly different impulses, i.e. that impulses are representational. But there is a law in the neurosciences, called the Law of the Specific Energy of nerves, that tells us, in Hayek's (1952, p. 10) formulation, that "the effect of the impulse is independent of the particular kind of stimulus which invokes it." In other words, whether the stimulation is

[7] Cf. Quine (1969a, p. 75): "The stimulation of his sensory receptors is all the evidence anybody has had to go on, ultimately, in arriving at his picture of the world." This does *not* imply the reducibility of theory to sensory events.

[8] I do not and, as I will show below, neither does Quine, think that there is any epistemological difference in kind between science and the result of many other forms of inquiry. The word 'science' should thus not elicit a scientistic reading. Naturalized epistemologists, as Quine (1969a) calls us, do not expect their account of inquiry to be any better than the very inquiry that is our object. Cf. Quine (1969a, p. 84): "we are after an understanding of science as an institution or process in the world, and we do not intend that understanding to be any better than the science which is its object." In other words, we are not trying to give a foundation to science that is any more secure than science itself.

to the retina, the skin on the right leg, or the mucous membrane of the nose, the impulses thence resulting are not inherently different.

But then how does the organism discriminate? According to Hayek, the organism distinguishes stimuli rather by the different topological location (p. 37) of the resulting impulses in the overall system of connections: a stimulus to the retina does not fire the same neurons as a stimulus to the nose. In a nutshell, our discrimination of stimuli is not the result of there being qualitative differences in the *impulses* they cause, but of "the position of the fiber in the central organization of the nervous system which carries the impulse." (p. 12)

This position is not static. As Hayek explains, the way the organism localizes the impulses caused by stimuli evolves with the organism's history, being dependent on previous occurrences and co-occurrences of the impulses. According to Hayek, the co-occurrence of impulses tends to connect the parts of the brain which they severally affect. From this, Hayek concludes that the topological location of the co-occurring impulses becomes more *similar*[9]. Since the organism discriminates stimuli by the topological location of the resulting impulses, the organism will tend to associate stimuli that cause co-occurring impulses.

In order to explain cognition at any given moment, Hayek introduces two notions, that of the map, and that of the model. The map is the semi-permanent (cf. pp. 114-5) classificatory apparatus which the individual has acquired throughout her history, whereas the model is the effect which the present situation is producing on the central nervous system. Naturally, whatever model is in force depends on the map from which it derives. Further, Hayek sees the map as being adjusted by taking into account the way the model fits not only the present, but also the *expected* sensory events. In fact, according to Hayek, speculation is of the essence: "the representations of the external environment which will guide behavior will thus be not only representations of the actually existing; but also representations of the changes to be expected in that environment. We must therefore conceive of the model as constantly trying out possible

[9] Hayek's (1952, p. 62) notion of similarity in this context is a three place relation: two *neurons* are equally similar to a third if they have the same *number* of connections. Yet, they may have no connections in common.

developments and determining action in the light of the consequences which from the representations of such actions would appear to follow from it." (p. 121)

For Hayek, the nervous system is thus engaged in a dynamic process of interpretation, of subsuming impulses under multiple classes. When its expectations are frustrated, the organism reclassifies: It changes *both* the map and the model. It makes adjustments to the topological location of specific impulses so that its classifications better "approximate" (p. 107) the relations between the stimuli affecting the organism. "The immediate effects of such conflicting experiences will be to introduce inconsistent elements into the model of the external world; and such inconsistencies can be eliminated only if what formerly were treated as elements of the same class are now treated as elements of different classes." (p. 169)

However, it is obvious that in order to *re*classify, the organism must first have classified. In Hayek's words, "an event of an entirely new kind which has never occurred before, and which sets up impulses which arrive in the brain for the first time, could not be perceived at all." (p. 142) In *The Sensory Order*, Hayek (1952, p. 168) explicitly recognizes that some connections are acquired "not by the individual but by the species." In general, it is impossible for any learning to occur without there being some prior discrimination that the learning organism could seize. As Quine (1973, p. 19) notices, "if an individual learns at all, differences in degree of similarity must be implicit in his learning pattern. Otherwise any response, if reinforced, would be conditioned equally and indiscriminately to any and every future episode, all these being equally similar. Some implicit standard, however provisional, for ordering our episodes as more or less similar must therefore antedate all learning, and be innate." [10]

It is the innate standards of similarity that form the first organization of sense experience. Depending on how the trade with the environment goes, the subject will need to reclassify: it will start noticing differences where none were, and identifying where before she distinguished. Thus, with obvious path dependence, new groupings

[10]An important lesson of experimental psychology is the lack of any need for explicit reinforcement in the process of learning. Cf. Putnam (1967, p. 114), cf. also Quine (1968, p. 31): "Like all conditioning, or induction, the process will depend ultimately also on one's own inborn propensity to find one stimulation qualitatively more akin to a second stimulation than to a third; otherwise there can never be any selective reinforcement and extinction of responses."

are formed, new saliences and focal points arise, and new standards of similarity are adopted.

An illustration is helpful at this juncture. Barring disabilities, it is common-sense that a newborn is capable of acquiring any language on the planet: a child born to French parents would speak Inuit if it were raised in an Inuit speaking community. This implies that newborns must be equipped with hardware up to the task. *Inter alia*, their vocal system must be capable of producing all the sets of phones found in languages around the world, and their hearing must not be deaf to the phonemic contrasts of the several languages.

Interestingly, newborns with just a few days of age can distinguish not only the language they have been surrounded by from other languages, but they are also capable of distinguishing between foreign languages to which they have never been exposed[11]. According to Guasti (2004, p. 30), this capacity results from "certain acoustic properties that capture infants' interest and that can be extracted from utterances in very little time and with limited exposure." It is the rhythmic properties of language that seem to be salient (p. 40). In fact, more than distinguishing languages, babies have been found to distinguish between classes of languages sharing rhythmic properties. For instance, babies have found it difficult to distinguish Dutch from English, two stress-timed languages, but not French from English, a syllable-timed and a stress-timed language. (p. 36)

More than this, newborns are capable of discerning phonological contrasts not only in the language that surrounds them, but also in foreign languages *that the adults in their community cannot discern with ease or at all* (cf. p. 42). Yet, if newborns are capable of discerning these many contrasts, at about the time they reach one year of age their discriminatory capacity is no greater than that of an adult in their linguistic community. In other words, the child becomes sensitive only to those contrasts that are important in the particular language to which they are continuously exposed (p. 43). This gearing of their standards towards what is relevant in their community is a gain in efficiency: it

[11] What to count as one language or two different languages is, of course, vague. But so is having a beard. Yet no one would deny that Marx was bearded and Joan Robinson was not.

makes the relevant phonemes focal and the child stops wasting valuable attention on irrelevancies.

But besides being able to discriminate sounds, as potential speakers, children must also be able to articulate them. An initial step in the acquisition of this skill is the phenomenon of babbling. It is through babbling that infants begin developing control over their vocal apparatus. They correlate the sounds they emit with particular vocal behaviour of theirs. The sounds emitted are, in turn, correlated with those they distinguish in their environment. Interestingly, if at first there is little difference in babbling behaviour across cultures, the time babbling becomes noticeably influenced by the native speech is also the time when children begin specializing their auditory discriminations. (pp. 49-50)

The lesson I wish to press is general. Where we notice differences, we separate, where we notice similarities, we associate. When our expectations are borne out, we reinforce them; when they are frustrated, we revise them: we change the distinctions we find important to make or to blur, we project along other paths, delineating new entities, and putting them in novel interrelations. When we are puzzled we offer bold redefinitions, say, calling uniform rectilinear motion ‘rest’, i.e. lack of change; we may hypostatize all sorts of exotic entities in the deepest parts of our theories in the hope that they will help us make sense of whatever troubles us, be they quarks or supernatural activity; and we may even revise the inferential relations we allow or the sources of evidence we are willing to condone. The process may be more or less explicit and articulate, of course, depending on whether we are babies first trying to make sense of the sounds we hear, or physicists trying to explain the recalcitrant results of an experiment.

With this general lesson in mind, in the next section I will take a closer look at our *particular* tendencies for discriminating, grouping and associating.

2.2. Induction

To group is to compare. To compare is to gauge similarities, and similarity is always dependent on a standard, along dimensions, and partaking of degrees. I find a red leather shoe more similar in color to an orange cotton shirt than to a blue woolen sock, but functionally it reminds me more of the latter. Finally, with respect to their material constitution, I find the shirt more similar to the sock than to the shoe. I expect most people with a background similar to mine to make the same judgments: both red and orange are warm colors, whereas blue is cold; socks and shoes are both worn by our feet, whereas a shirt dresses our torso; and wool and cotton are both fabrics, whereas leather is tanned skin. But, of course, we can easily come up with different standards for comparing, even along the same dimensions. Why not team blue with red? They both feature in the Union Jack, whereas orange is the color of the Dutch royal family. Socks are also functionally very much like shirts: both are in direct contact with the skin and you do not take them off when you arrive home. Finally, wool and leather are both animal produce; not so with cotton.

Other dimensions of comparison may easily be conjured up. Some dimensions are more often used than others, more salient or focal than others, and more likely to produce shared standards among people in a community than others.

In any event, we *do* compare, and therefore group. Moreover, when we notice that elements in one group tend to also be elements of another group, i.e., when we find a correlation between two dimensions, we tend to project. If we notice that whenever there is a similarity of such-and-such, there is also similarity of so-and-so, when we see a such-and-such we often expect to see a so-and-so. We begin noticing regularities between groupings of our own making. When Jane takes the bus, she arrives late to class. When interest rates go up, asset prices go down.

Donald Davidson (1995a, p. 215 italics supplied) writes that "long prior to the acquisition of language, or of anything that can properly be considered concept formation, we act as if we had learned crude laws. *We are inducers from birth.* (…) In the course of avoiding and seeking, learning to control our environment, failing and

succeeding, we build the lawlike *habits* that promote survival and enhance life. These laws of action are highly pragmatic not only in their conspicuous ties to action, but in their breezy disregard of the irrelevant or implausible."[12] But the question begs to be posed: *how* do we project? It seems that we have landed straight into the jaws of induction and its riddles, old and new[13].

Imagine you see an object with certain conspicuous features that you associate with precious stones. Someone tells you that the object is indeed a precious stone, in fact an emerald. You notice that it is green. Every time someone tells you that you are in the presence of an emerald, you notice that it is green. Sooner or later, I suppose you come to expect that all or most emeralds are green. But what about *grue*? Something is grue if observed before January 1st 2100 and green, and blue if *not* observed before January 1st 2100. Clearly your experiences with emeralds are as much positive instances of "All emeralds are green" as of "All emeralds are grue." Why do you not instead come to expect that they are *grue?* It *seems* foolish to project along grue. But *why*?

Quine (1969b) associates this issue with Hempel's famous problem of black ravens. Take the statement "All ravens are black". Suppose you observe a green leaf, i.e., something that is not black and also not a raven. Why does this observation not lead you to expect that everything that is not black is not a raven, which is equivalent to saying that all ravens are black[14]? Again, why do we not project along these paths? Why do we project green of emeralds, but not grue, why do we project black of ravens, but not non-ravenhood of non-blackness?[15]

In keeping with the naturalistic perspective I adopted earlier, the authors I have been working with, viz. Quine, Goodman and Davidson, are more interested in studying our

[12] Cf. Hayek (1952, p. 130): "It is conceivable that a structure endowed with the capacity of retaining experienced connexions [sic] might learn separately the appropriate responses to most of the possible combinations of events. But if it had to cope with the complexity of its environment solely by classifying individual events and learning separately for every combination of such events how to respond, both the complexity of the model required and the time needed for building it up would be so great that the extent to which any given structure could learn to adapt itself to varying circumstances would be very limited."

[13] The New Riddle of Induction was first so called by Goodman (1983). The presentation in the following paragraphs is based on his work. An economist who has dedicated efforts to these issues is Gilboa (2009).

[14] $\forall x(Rx \rightarrow Bx) \Leftrightarrow \forall x(\neg Bx \rightarrow \neg Rx)$, where Bx means 'x is black' and Rx means 'x is a raven'.

[15] Davidson (1995a, p. 207) summarizes the riddle nicely when he writes that "similarity cannot, by itself, carry the burden of distinguishing the lawlike from the non-lawlike."

inductive *practices* than in justifying them[16]. According to Goodman (1983), the difference between 'green' and 'grue' is that the former is much more *entrenched* than the latter. A predicate is more entrenched than another if, in the course of linguistic use, it is more often projected than the other. Actually, Goodman's position is more nuanced than this. It is not really *predicates* that are more or less entrenched, but the classes to which they *refer*, i.e. their *extensions* or *denotations*[17]. A class of things is more entrenched if the predicates that refer to it are habitually projected[18].

Still, the crucial point in Goodman's position is his tying up entrenchment with our projecting *practices*, as exercised in our use of language. What is salient to us or in what directions we tend to project integrates the wisdom that has been accumulated by our communities, and is handed down to us when we learn to think and to speak. As we have seen, in our dynamic process of making sense of the world, we discriminate and identify along one or another salient dimension. We *also* experiment with projections. If those associations make us happy, we keep doing them, and with time the relevant

[16] Cf. Quine (1969b, p. 127): "For me then the problem of induction is a problem about the world: a problem of how we, as we now are (by our present scientific lights), in a world we never made, should stand better than random or coin-tossing chances of coming out right when we predict by inductions which are based on our innate, scientifically unjustified similarity standard. Darwin's natural selection is a plausible partial explanation." Goodman (1983, p. 64) writes: "Predictions are justified if they conform to valid canons of induction; and the canons are valid if they accurately codify accepted inductive practices." Davidson (1991, p. 194) writes: "I do find congenial Quine's resolutely third-person approach to epistemology, and to the extent that the naturalization of epistemology encourages or embraces such an approach, I am happy to count myself a naturalized epistemologist."

[17] The reference/extension/denotation of the name 'Hayek' is Hayek, i.e. the only Austrian winner of the Nobel Prize in Economics

[18] Goodman (1983, p. 95) writes that "the entrenchment of a predicate results from the actual projection not merely of that predicate alone but also of all predicates coextensive with it. In a sense, not the word itself but the class it selects is what becomes entrenched, and to speak of entrenchment of a predicate is to speak elliptically of the entrenchment of the extension of that predicate. On the other hand, the class becomes entrenched only through the projection of predicates selecting it; entrenchment derives from the use of language." Goodman's position has been criticized by Davidson who objected to Goodman on an important point. Davidson does not believe that classes are entrenched *tout court*, but that we should look at induction by way of a *two-place* relation between classes. Cf. Davidson (1995a, p. 208): "For him [Goodman] projectibility is a property of predicates, not of predicates relative to other predicates." (Davidson's speaking of predicates as opposed to their referents should be taken as careless, elliptical phrasing.) Davidson (1970, p. 218 italics supplied) thus prefers to say that "grueness is not an inductive property *of emeralds*". But it *is* an inductive property of something, for instance of *emerires*. Something is an emerire if it is observed before January 1st 2100 and is an emerald, otherwise is a sapphire. Cf. Davidson (1966). Sure, we may not find the notion of grueness or that of emerire very useful - indeed the classes to which they refer are not entrenched. But Davidson (1970, p. 218) writes that "we know a priori [that predicates] are made for each other – know, that is, independently of knowing whether the evidence supports a connection between them. 'Blue', 'red', and 'green' are made for emeralds, sapphires, and roses; 'grue', bleen', and 'gred' are made for sapphalds, emerires, and emeroses." We should not make any mystery of Davidson's 'a priori' in this passage. It just means that there are classes that go well with each other according to our projective habits, even if they do not fit.

classes become entrenched[19]. Why greenness and not grueness? Simply because greenness became more salient, and has served us well on numerous occasions when speaking of emeralds, and other things[20].

2.3. Language-games and Vocabularies

I started this chapter by noticing that there are many ways of worldmaking. As I mentioned, not all of them are interesting, and some conflict. But it also seems obvious that worldversions not only need not conflict, but often complement one another. Indeed, the vision expressed by a painting is not competing in interest with an article in the latest issue of Nature, and the warmth, affection and wisdom of an old grandmother's advice on getting rid of a bad cold loses none of its importance because it would make poor WHO recommendations.

Wittgenstein ([1953] 2010, §23, pp. 14-15 italics in the original) introduces the expression "language-game" (Sprachspiel) to "emphasize the fact that the *speaking* of language is part of an activity, or of a form of life." The painting and the article, the grandmother's piece of advice and WHO recommendations, belong to different games, if not different language-games. We cannot understand the significance, interest and importance of each in a vacuum, *absolutely*, the same way that we cannot understand what people mean if we take their sentences in a vacuum, without relating them to the purposes of the utterer and contextualizing them in the circumstances of utterance.

An example will clarify. Suppose you overhear an economist lecturing her son. She is explaining to him that if he does not want to be a bum, working odd-jobs and living hand-to-mouth, he must study hard, have a moral compass, respect his friends and associates, etc. Later that day you attend one of that economist's lectures, where she relates the figures of unemployment to the recessive state of world-markets, she argues that poor decisions of policy have caused a mismatch between skills supplied and skills

[19] Some care should be exercised here: Cf.Goodman (1983, p. 97) "Again, a very familiar predicate may be rather poorly entrenched, since entrenchment depends upon frequency of projection rather than upon mere frequency of use."
[20] Naturally our inductive practices also depend on our *personal* history, and on where we are in life.

demanded, and she claims, ironically, that her students' bourgeois moral sentiments are the ineluctable side-effect of class struggle. Would you find her inconsistent?

On the one occasion she is trying to turn her own, very concrete, child into a person she can be proud of, on the other she is trying to retain the attention of a bored audience and motivate it to spend time thinking about important abstract matters. She is doing different things, participating in different activities, playing different (language-)games. As a competent social being she adjusts her ways, more or less radically, to the game she is playing. In the first game she uses loose words, charged with individualistic moral significance, and she makes use of her parental powers over the child to discipline him. In the second game she tries to sound scholarly and to be precise, and she makes use of the hierarchical relation between teacher and students, if not of the intellectual fascination she exerts over them, in order to supply them with food for thought.

Used for a variety of purposes, under variegated circumstances, articulating many worldviews, our natural language is a hodgepodge of *vocabularies*[21]. Wittgenstein ([1953] 2010, §18, p. 11) expresses this with a noteworthy simile: "our language can be regarded as an ancient city: a maze of little streets and squares, of old and new houses, of houses with extensions from various periods, and all this surrounded by a multitude of new suburbs with straight and regular streets and uniform houses."[22]

I find this simile masterful because you can hardly press it too hard. As a city is ancient, our language is as old as we are[23]. As a city evolves, so does our language: vocabularies appear and evolve in tandem with social interaction, and social phenomena, as new

[21] As with games, it is through social conditioning that we learn the several vocabularies available, how to use and mix them together, and also the right way to put things in this or that situation. There are no necessary or sufficient conditions telling us where one language game ends and another begins, or where one vocabulary makes sense and another does not. It all comes down to judgment, and to the grasping of the contextually relevant similarities and differences.

[22] A vocabulary could be seen as a street, sometimes merging almost imperceptibly with another street, other times forming well-delineated avenues; instead of streets, some vocabularies are more like squares, concentrating traffic, with streets radiating it; others may even best be seen as a manor house, opening its doors to a small elite alone, but commanding the deferred respect of the excluded peasant. Vocabularies are also neighborhoods, sometimes modern, geometrical, spacious, and well-organized, as the vocabulary used to do theoretical physics; other times they are old, with streets crossing at random, with dark alleys and cul-de-sacs, yet full of secrets, inspirational detail, and charm – somewhat as our everyday explanations of why, say, Jane fell in love with Edgar during that May weekend.

[23] Or maybe we are as old as our language.

groups of people do new things, have new ideas, devise new applications, and, importantly, do so with ever transformed worldversions.

Vocabularies not only adjust to pre-existing situations and games, but their appearance, development and application also (re)creates. As in cities there are derelict neighborhoods, with the transformation of worldviews there are language games we stop playing, and vocabularies which we end up, for a while at least, abandoning. Together with changed games and vocabularies, we find changed theories. We no longer say that sickness is the result of unbalanced humors, *dyscrasia,* or that bad crops result from the Wrath of God. We have developed other ways of defining and explaining illnesses and other predicaments, ways that we believe to be better: more truthful, perhaps, and offering us better chances of achieving this or that goal. Other times, instead of abandoning, we *remodel* our neighborhoods. Similarly with vocabularies, games and theories. We now see alcoholism as often not just the result of poor choices and gluttony, but also as a disease, and we thus expand our explanation and understanding of alcoholism by using the vocabulary and theories of psychiatry to make sense of and deal with its most formidable manifestations.

The fact that vocabularies are adjusted to their language games, and vice-versa, helps explain why most of our ordinary, everyday notions are vague, open in application, and in a loose relation to each other[24]. These everyday vocabularies are loose and flexible enough to be of service at all the required uses, in all typical circumstances; and yet, they are still tight and strict enough to be more than random sounds or shapes. There is something to be gained if everyday vocabularies offer people the means to make sense of the reasonable, the determined, the probable, but also allow them to cope with the arbitrary, the erratic, the unheard of.

On the other hand, their flexibility and looseness usually make them poor at specialized games, for instance, when we wish to develop careful explanations, or to erect a strict calculus. They also offer poor grounds for finding strict relations between notions, or to

[24] That they are so can be illustrated by trying to define "friend", and relate it to, say, the notion of "family" – you are always bound to find an intuitive counterexample against any strict definition, or any strict relation between the two.

make strict predictions. These are not their games. Other vocabularies are better for these purposes. As shown in 2.1., as people try to develop new, possibly deeper explanations of whatever troubles them, they learn to talk in new ways, to follow different inferential paths, to posit new entities and to get rid of old ones. They may become more explicit about the rules of inference they would like to follow, and more demanding with the forms of evidence that they are willing to accept. Historically, it is true that people often start with refinements of notions that their everyday vocabularies give them, say the ordinary notion of price, in order to explain phenomena that are troubling to any thinking human, say, why the price of corn and other grains went up last year. With time and leisure, they frequently end up replacing those everyday notions with others, often homonymous, that hopefully help them arrive at persuasive answers to their changing questions. The ordinary notion of 'price' transmutes into its homonymous 'price' as used in modern economics, and is now used not only to study Corn Laws in 19th century England, but also to answer highly abstract and technical questions, such as whether the interest rate of a stationary economy is zero.

Inquiry is the human activity of looking for knowledge and wisdom. Inquiry in general, and scientific theories in particular, are a precipitate of the interaction of groups of people under all sorts of institutional arrangements, applying all kinds of vocabulary, and vying for all the things that humans vie for: knowledge, prediction, control, social approbation, wealth, spiritual salvation, etc. Still, this sociological dimension does not impugn the interest of epistemological reflections about the types of explanation that one person or another may, or should, strive for, and what their reasons are, when we take into consideration the sort of things that they take upon themselves to explain, predict, or control, and what they want their knowledge for: to achieve aesthetic bliss, to implement social policy, to satisfy curiosity, to build a machine, to win an argument, etc.

2.4. The World Well Lost[25]

I started this chapter by agreeing with Goodman's observation that there are many worldversions. Goodman (1978)'s "adversary is the monopolistic materialist or physicalist who maintains that one system, physics, is preeminent and all-inclusive, such that every other version must eventually be reduced to it or rejected as false or meaningless" (p. 4). Perhaps his adversary is the gentleman who told me that constellations do not exist.

In the last three sections I tried to paint a picture of what such versions are, their genesis and social contexts. Still, beyond stating that there could be many interesting world-versions, Goodman's point is that "versions can be treated as our worlds" (p. 4), and that versions, and hence worlds, are of our *making.*

But there seems to be something suspect about this position. After all, we may very well agree that there are many worldversions, many right ways of describing this or that, and of explaining and accounting for whatever puzzles us. We may further agree that those versions sometimes are irreducible to one another, that they are useful for different purposes, and that they make sense in different games. But this agreement seems to presuppose that versions must be versions *of* something, this something existing independently of our versions of it - an underlying World, perhaps. In this vein, a truthful worldversion would be a version that faithfully represented this World, even if only a particular dimension thereof.

Yet, Goodman is quite explicit in telling us that he has no interest in *the* one World, *underlying* all versions, and existing ready-made outside all descriptions (Cf. Goodman (1978, p. 4))[26]. The problem Goodman sees in such an underlying World of pure content, the common-denominator of all the right versions, is that there is nothing we can say about it. As he (p. 6) writes, "content vanishes without form. We can have words without a world but no world without words or other symbols." Whatever we say

[25] I borrow this title from Rorty (1972).
[26] As Quine (1978, p. 96) writes, "I cannot quite say versions of the world, for Goodman holds that there is no one world for them to be versions of. He would sooner settle for the versions and let the world or worlds go by."

about such world is already a worldversion, a version of our making, to be assessed with recourse to other versions we have available. "We cannot test a version by comparing it with a world undescribed, undepicted, unperceived", he writes (p. 4). And so, "we are confined to ways of describing whatever is described" (p. 3).

In a sense, Goodman is emphasizing what Lawson (1997, p. 25) calls the transitive dimension to knowledge. As the latter puts it (p. 25 italics supplied),

"It is necessary (…) to recognize a *transitive* dimension to knowledge. (…)That is, it is necessary to recognize a dimension of *transitive* objects of knowledge, including facts, observations, theories, hypotheses, guesses, hunches, intuitions, speculations, anomalies, etc., which condition all further knowledge. (…) Knowledge must be recognized as a produced means of production (of further knowledge) and science as an ongoing transformative social activity."

In a similar vein, Goodman (p. 6) writes that "worldmaking as we know it always starts from worlds already on hand; the making is a remaking." To say that there is a transitive dimension to knowledge is to recognize that knowing is, ineluctably, a human affair. Goodman is telling us that we can never transcend *every* standpoint; that to judge, or even to create a new worldversion, we must always utilize other, pre-existing versions.

But, besides a transitive dimension to knowledge there is also, of course, an intransitive dimension. *Given that* the sentence 'snow is white' as uttered by me means that snow is white, then it is true if and only if snow is white, whatever my attitudes towards this sentence are. Whether I, you, or anyone else *believe* it true or false is, obviously, irrelevant to its truth or falsehood[27].

[27] When we speak of the intransitive and of truth, some people talk of *correspondence*. We could, of course, say that a sentence is true if and only if it corresponds to the facts, or some other such roundabout way, but this would only go beyond saying that the sentence is true if we could spell out the relation of correspondence and what the facts *are.* As Davidson (2009, p. 41) writes, "the correct objection to correspondence theories is (…) that such theories fail to provide entities to which truth vehicles (…) can be said to correspond." It is not uncommon for people to pay lip service to "correspondence" simply because it seems to crystalize the non-epistemic or intransitive dimensions of knowledge and truth. Davidson (1969) is a case in point. His views on truth suffered no radical change, but in his later work he abandoned any notion of 'correspondence' as misleading. To say that sentence 'X' is true because it corresponds to the facts is just to say that sentence 'X' is true. And sentence 'X' is true iff X.

By emphasizing world*making*, Goodman is not denying an intransitive dimension. As he (p. 91) writes, "we must distinguish falsehood and fiction from truth and fact." Rather, he is drawing attention to how much the transitive imposes on the intransitive: we may talk, associate, and induct in all the ways we have seen above. We may speak of green emeralds, grue emerires, featherless bipeds, Germans, Nobel Prize winners, aggregate demand, or space-time. The distinction between truth and falsehood cannot thus be done "on the ground that fiction is fabricated and fact found" (p. 91). As Lawson (1997, p. 59) puts it, "knowledge, rather, exists in a historically specific, symbolically mediated and expressed, practice-dependent, form."

Naturally, as Goodman (p. 95) writes, "though we make worlds by making versions, we no more make a world by putting symbols together at random than a carpenter makes a chair by putting pieces of wood together at random." Indeed, the carpenter who wishes to make a chair must not only put pieces of wood together in thought-out, non-random ways, but he must also judiciously select the instruments and techniques he will use to work his raw materials. As Lawson (1997, p. xii) urges, and we often forget, the same is true of science and worldmaking in general: our methods should be judiciously chosen and adapted to what we are trying to know and do.

Moreover, as Goodman emphasizes (p. 21), "a broad mind is no substitute for hard work." Quite the opposite: the person who agrees with Goodman that there are many complementary worldversions faces a much harder task than the monist or physicalist who rejects pluralism right off the bat, or than the postmodernist who a-critically accepts any worldversion. Wise folk have to ponder, and adjudicate. They must also study the interconnections between versions, and must organize their pluralism into a coherent whole that makes it possible to apply the right version, in the appropriate games.

It is thus clear that from the fact that worlds are of our making, and that there are many truthful worldversions of independent interest, it does not follow that there is no objectivity or intransitivity. Moreover, since inquiry answers to such a variety of goals, and is conducted under so many social and psychological contexts, the necessity of a judicious assessment of the right way, or the right method, of conducting it is ineluctable.

2.5. Summary

In this chapter I argued that human beings worldmake, that they develop plural worldversions, vocabularies, and theories to make sense of their environment, to do all sorts of things, and to reach all sorts of goals. I emphasized that to know is a human affair, and discussed some of the psychological and social determinants of worldmaking. I also argued that from the fact that we see things in *our*, human, way, it does not follow that truth is not objective, or that there is no intransitive dimension to knowledge.

This chapter thus offers a reflection on a fundamental matter to the moral scientist, since, as Hayek argues in his *Scientism* essay, action cannot be understood without knowledge of the way agents make sense of the world.

3. The nature of inquiry in Hayek's 'Scientism and the Study of Society'

Hayek's (1942), (1943b) and (1944) three-part *Scientism and the Study of Society* is a remarkable piece: its major flaw is also its greatest virtue – it is, at times, quite opaque. This opaqueness results in frustration, doubt, and potential incoherence, and *therefore* has spurred an extraordinary number of interpretations, many of which are interesting contributions[1]. In this three part article, Hayek offers methodological reflections on inquiry, both in the natural and the moral sciences. I am mostly interested in Hayek's positions on the moral sciences. But since these are incomprehensible without first going through what he says about the natural sciences, I will examine these in detail. In particular, I will seize specific passages of Hayek's, and try to derive their logical implications, often going further than Hayek does in this respect.

I will try to turn the above noted opaqueness into opportunities for the elucidation of difficult points of economic methodology. When needed, I will make use of other works of his written at about the time of this article.

3.1. Classification and Reclassification

In the *Scientism* essay, and anticipating some of the views in *The Sensory Order* which I presented in 2.1., Hayek views natural science as a dynamic process of classification and reclassification, of revision and reconstruction of "the concepts formed from ordinary experience" (p. 82). Hayek describes scientific endeavors as issuing from our dissatisfaction with existing common-sense explanations of phenomena, and as attempts

[1] Cf. Caldwell (2005, Appendix D) who writes that this essay is a Rorschach Test, saying more about the interpreter than of Hayek. Among the numerous interpretations, we have that of Burczak (1994) who sees it as postmodern, Lawson (1997, Ch. 10) who sees it as almost positivistic, Caldwell (1994) who sees it as as anti-modernist and non-hermeneutic, Madison (1991) who sees it as hermeneutic, and Runde (2001) who sees in it an ontologically sound piece.

to replace our pre-scientific conceptualizations with a theoretical framework "based on consciously established relations between classes of *events*" (p. 84, italics supplied). Hopefully, such a framework will supply us with "general explanations" and "general rules" (p. 84) about the behavior of phenomena. (Cf. Hayek (1952 p. 3))

According to Hayek,

A) "[Science] *begins* with the realization that *things which appear* to us the same do not always behave in the same manner, and that things which appear different to us sometimes prove in all other respects to behave in the same way; and it proceeds from this experience to substitute for the classification of *events* which our *senses* provide a new one which groups together not *what* appears alike but what *proves to behave* in the same manner in similar circumstances." Hayek (1942, p. 83 italics supplied)[2]

Interestingly, Hayek seems to take his "science begins" rather literally[3]. He writes that modern Science began in the Renaissance (p. 81), and that its "ways of thinking" (p. 81) had to "fight their way" (p. 81) against the established, pre-scientific frames of mind. The latter were anthropomorphic and animistic in character, mostly limited to the study of ideas, either those of fellow men, or God's. These scholastic ways of thinking were replaced by a scientific attempt to "get down to 'objective facts'" (p. 82).

3.2. What is Science?

There is thus, in the *Scientism* essay, a conception of natural science as a particular *kind* of inquiry, and therefore one that we should be able to *tell apart* from other kinds. The

[2] The italics are an example of Hayek's somewhat careless phrasing. Do we classify things or events? Or both? What ontology is here implicit? What is the "behavior" of things supposed to be? Is it what they cause? Do *things* cause anything? Or do they have dispositions? If so, at what level should we recognize this? Actual? Counterfactual? Transfactual? What is the relation between "things" and "events"?

[3] In many places, Hayek explicitly rejects foundationalism. He has, however, been accused by Lawson (1997, p. 194) of committing this sin in the *Scientism*: "Hayek's achievement in his scientism essay is not so much a transcendence of positivism and its errors as a sideways shift of it all; a move towards a subjectivised version. In place of the 'brute facts' of positivism we find, in effect, the brute opinions, beliefs and attitudes of hermeneutical foundationalism."

general remarks found in A), however, are insufficient for this. Although in A) Hayek explicitly refers to science, we have seen in chapter 2 that everybody, as they go about their business and attempting to make sense of the passing show, transform the concepts and explanations they have come into during their history into others that they find more plausible, or more pertinent for this or that purpose. This may be done more or less consciously, and more or less systematically, using whatever vocabulary makes better sense in the language game people are concerned with. The difference from what scientists do, however, seems, pending further argument, to be one of degree.

3.2.1. Telling Natural Science Apart

Indeed, it is easy to think of an everyday example of the sort of 'classification' that Hayek mentions in A). Imagine your colleague Jane complains that the door of her car keeps getting scratched. She is not sure whether the scratches happen at night, when the car is parked outside her home, or during the day, when it is parked outside your office. You offer consolation and a few hypotheses: maybe she keeps bumping into something without noticing, or maybe she has a truculent neighbor… A few days later, Jane tells you that she stopped parking her car in parking spots that have a slant. She believes that the doors of cars parked in such spots tend to get scratched.

There are many possible reasons why Jane came to hold this belief: maybe she came to notice that her car got scratched only when she parked it in certain spots, which she then noticed to have significant slant; maybe she observed another car getting scratched after its door was opened against the floor; or maybe she is given to superstition and an Apparition told her so. Whatever Jane's reasons, whether we find them good or bad, what she did is clear: she classified a bunch of entities together, those which satisfy the formula "*x* is a slanted parking spot", and consciously put them in an inductive relation with another bunch of entities, those satisfying "*x* is a scratching of a car". She defined the two classes of entities by using observable characteristics, viz. slant, scratching, and the quality of being a car, as defining features. In fact, she grouped things in a rather ingenious way: the slant of parking spots is not, at least to me, a particularly salient

feature of spots. Jane thus obtained a moderately general sentence – "if the spot is slanted, then cars parked are likely to get scratched" - which she finds true and gives a sufficiently good reason for her to avoid parking in those spots. Naturally, her belief may be right or wrong, and its truth value is independent of her believing it. If her car keeps getting scratched, clearly she misidentified the cause.

I believe that few would find Jane's theory scientific in any interesting sense[4]. But not only does Jane's procedure observe the criteria in A), her theory is of some generality. Certainly it is not less general than most generalizations we aspire to in natural sciences like meteorology. There thus only seems to be a difference of degree along the lines of A) between 'science' and other more local and less articulated theories that make sense of concretes of time and place, or have reduced inductive applicability[5].

Perhaps the author of the *Scientism* essay would agree with me. Maybe he does not think that, at present, there are any synchronic differences in kind between science and other ways of worldmaking. Perhaps the overall conception of inquiry I presented is itself peculiar to our scientific age, and thus Hayek is only committed to there being diachronic differences between the excessively teleological explanations of a pre-scientific era and those we now offer, whether as scientists or as ordinary folk. Indeed, Hayek writes that "we live now in an atmosphere where the concepts and habits of thought of everyday life are to a high degree influenced by the ways of thinking of Science." (p. 81)

Still, I doubt that this effect of what Hayek calls 'scientific ways of thinking' on common-sense is to the degree that would be necessary to support his position. It may be true, although I do not believe we have good reasons to think it is, that people are now less persuaded by teleological arguments of the kind that explain crop failures by

[4] Hayek (1942, p. 87 italics supplied) writes that "the picture which man has actually formed of the world and which guides him well enough in his daily life, his perceptions and concepts, are for Science not an object of study but an imperfect instrument *to be improved*." - improved for Science's purposes, of course. We can hardly expect to use quantum mechanics to fix a water leakage: it is useless for these purposes, not only as a matter of computational difficulty but as a matter of principle. Nor does it seem that "science" could help Jane that much… My point is that we do not have a static and well-defined picture of the world. People, in their daily life, learn, forget, change, etc. This process does not differ from what the scientist does along the lines of A).

[5] Cf. Quine (1969b, p. 129): "Sciences, after all, differs [sic] from common sense only in degree of methodological sophistication."

the Wrath of God, or an indigestion by the evil eye of envious in-laws[6]. But it is difficult to believe that our skills at judicious classification and reclassification have dramatically changed with the appearance of modern science. The people who carefully selected the specimens they bred to achieve the most desirable characteristics in their domesticated creatures, or the ancient Egyptian who put moldy bread on open wounds to fight infection were certainly guided by judicious inductions at fortunate circumstances.

In any event, this is not an argument that Hayek can safely adopt. When we get to discuss his reflections on the moral sciences, we will see that Hayek takes the intersubjective similarity of our classifications as a necessary condition for social science. But then, if he followed the argument we are considering to the point that seems necessary, he would be saddled with the implausible position that we cannot do moral science about cultures untainted by western science, or our own culture before the Renaissance. There is abundant evidence that Hayek by no means believes this. So we should waive this interpretation aside.

3.2.2. Reception and Perception

Maybe Hayek does not think that the *differentia specifica* of natural science lies in the general reclassification of concepts. In fact, Hayek seems to believe that the sciences reclassify not only, as A) suggests, classes of complex entities such as macroscopic bodies like cars, or events like scratchings, but the very sense *qualities*:

[6] Whatever we make of these explanations, it seems to me that they belong to different games than scientific explanations. A farmer may well believe that her crops have failed because she was being punished by God, *and* that God's way of doing this was by making the weather bad. None of this, of course, means that our common-sense ways of thinking are not impacted by scientific fads. Cf. Putnam (1975a, p. 358): "It seems clear that plausibility and probability have something to do with the accepted science and metaphysics of a given time. Teleological explanations seem plausible to an age that is steeped in teleological philosophy; mechanistic explanations will seem plausible to an age that is steeped in mechanistic philosophy."

B)

1. It may, however, still sound surprising that what is true of these provisional abstractions [mentioned in A)] should also be true of *the very sense qualities which most of us are inclined to regard as the ultimate reality*. Although it is less familiar that science breaks up and replaces the system of classification which our sense *qualities* represent, yet this is precisely what Science does. Hayek (1942, p. 83 italics supplied)
2. While the naïve mind tends to assume that external *events* which our senses register in the same or in a different manner must be similar or different in more respects than merely in the way in which they *affect our senses*, the systematic testing of Science shows that this is frequently not true. p. 83 (italics supplied)
3. This process of reclassifying '*objects*' which our *senses* have already classified in one way, of substituting for the 'secondary qualities' in which our senses arrange external *stimuli* a new classification based on *consciously* established relations between classes of *events* is, perhaps, *the most characteristic aspect of the procedure of the natural sciences*. The whole history of modern Science proves to be a process of progressive emancipation from our *innate classification of the external stimuli* till in the end they completely disappear. p. 84 (italics supplied)[7] [8]

[7] A third passage that would be relevant at this juncture is: "In the course of this process [that of reclassification by Science] not only the provisional classification which the commonly used *concepts* provided but also *the first distinctions* between the different *perceptions* which our senses convey to us, had to give way to a *completely new and different* way in which we learned to order or classify the *events* of the external world" p. 82 italics supplied.

[8] Notice how breezily Hayek runs the terminological gamut in these passages: sense qualities, external stimuli, objects and events, without ever giving criteria for their individuation. No wonder this essay has been the source of so many interpretations. In Hayek (1952, p. 2) he writes "we shall employ the term sensory 'qualities' to refer to all the different attributes or dimensions with regard to which we differentiate in our responses to different stimuli. We shall thus use this term in a wide sense in which it includes not only quality in the sense in which it is contrasted with intensity, extensity, clearness, etc., but in a sense in which it includes all these other attributes of a sensation." In Hayek (1943b, p. 111f italics supplied) he is somewhat apologetic: "[that all mental phenomena, sense perceptions and images as well as the more abstract 'concepts' and 'ideas', must be regarded as acts of classification performed by the brain] must also serve as a justification for what may have seemed the very loose way in which we have throughout, in *illustrative enumerations* of mental entities, indiscriminately lumped together such concepts as sensation, perceptions, concepts, and ideas."

What can we say about this (re)classification that is the "most characteristic aspect of the procedure of the natural sciences", as written in B.3.)? Apparently, it is the replacement of sensorial classifications, based on the effects events have on us, with *consciously* thought out classifications that transcend the 'secondary qualities' of events and capture the relations of similarity and difference that they really have.

In B.2.) Hayek goes so far as to tell us that there may be two events completely equal or different except in the way they affect our senses. In order to make his position on this clearer, I propose to make a distinction between *reception* and *perception*. Receptions are effects of external events on our sensory *surfaces*, whereas perceptions are the configurations of sensory qualities thence resulting. I assert that if there are differences in reception, there must be some difference in what is *causing* the reception. This is a trivial consequence of the principle 'same cause, same effect'. But what about the relation between perceptions and receptions? Could we have different perceptions without different receptions, or vice-versa?

Yes. It is not hard to find plausible examples of this. If we have just had something very sweet, then an orange will taste sourer than it otherwise would. We would say that the difference is not in the orange but in us. To understand what this means we have to look at its language-game. If someone asked us if the oranges are sweet because they would fancy one, we would say "I did not find them so. But it might have been me: I have been eating chocolate." Still, the *only* way that the same pattern of stimulation could lead to different perceptions would be if the sensory organism were in *different internal states* at the time of identical receptions.

But then the difference has only been shifted to a difference in internal states. These differences are *also* the province of the natural sciences. Clearly, the possibility of two events being different in no *other* respect than their effect on humans defies credulity. It would entail, *inter alia*, that physical science, which is used by Hayek (1942, p. 84) as an example of a science that has transcended our senses to a great extent, would have to take a notion like "sentient body" as primitive! How else would it be able to

differentiate events that have absolutely no *other* difference than in the perceptions, i.e. non-physical states, they cause?[9]

So, clearly, there cannot be events whose *sole* differences are on their *effects* on our senses. They must have *other* differences than these. This points to a less strict possibility, where we attribute to Hayek the position that there are *some* differences that are not relevant *for science's purposes;* and that *other* differences, to which we are, on a daily basis, blind, are important for the sciences. I have already shown in 2.3. that similarity is always along a dimension, relative to a standard, and partaking of degrees.

Just after B.2.) Hayek gives the following example:

C) "A white powder with a certain weight and 'feel' and without taste or smell may prove to be any one of a number of different things according as it appears in different circumstances or after different combinations of other phenomena, or as it produces different results if combined in certain ways with other things." Hayek (1942, p. 83)

Notice that whereas in B.2.) Hayek talks of discriminating between *events* that are registered by our senses in the same way, in C) he gives an example, supposedly illustrating what is said in B.2.), to the effect that some *things* may sometimes be indistinguishable to the senses, yet in other circumstances are associated with different, obviously *perceptual,* effects.

Notice also that *powders* are a rather theoretical notion: the class of powders is certainly *not* reducible to sense qualities and so, clearly, all this is just an elaboration of what he had said in A). There, I interpreted him as saying that sometimes we have to refine our notions to deal with new, recalcitrant situations: we have ever new interactions with the world, and these make us revise our conceptualizations. Maybe until recently those white powders were all believed to be the same, but now someone has been led to conclude, perhaps the hard way, but certainly by way of sense impressions that *are classified the same way they used to*, that some are inflammable and the others not. So

[9] As Quine (1978, p. 98) writes: "nothing happens in the world, not the flutter of an eyelid, not the flicker of a thought, without some redistribution of microphysical states. (…) If the physicist suspected there was any event that did not consist in a redistribution of the elementary states allowed for by his physical theory, he would seek a way of supplementing his theory."

the first are henceforth named "inflammable white powder" and the others "non-inflammable white powder", and we can derive the true law "if an inflammable white powder is put into contact with a source of combustion, it will burn." Hopefully other distinguishing criteria beyond inflammability might be found so that we can develop statements that are not trivial.[10]

In fact, Hayek (1952, p. 9 italics supplied) will write that "the distinction between different stimuli (...) must be independent of the different effects they have on the organism. *This independence can never be complete, since all our knowledge is derived from our sensory experience*. But it can be independent in the sense that we can classify the stimuli not according to their direct effects on our senses, but according to the effects which they exercise on other external events, *which in turn act as stimuli on our senses*." Exactly! Reclassification involves exposition to *further* patterns of stimulation, perhaps under particular circumstances, as in an experimental situation; it does *not* involve a reclassification of "the very sense *qualities* which most of us are inclined to regard as the ultimate reality."

I draw the reader's attention to the Lockean smack in Hayek's 'secondary qualities' (B.3.), which points in the direction of a distinction between scheme and content, between the properties that things or events have in themselves, their 'primary qualities', and those that our senses represent *them* to have. Perhaps science gives us the former, and so *that* is what distinguishes science from other forms of inquiry. After chapter 2, it should be clear that this distinction is problematic at best, with little to no pragmatic import in any event. If Hayek is explicit in seeing science as a human affair, without any source of evidence besides that provided by sensory events, there is also in Hayek an insouciant use of pronouns, and of (other) implicit assumptions that point to an independently structured world which knowledge tries to *represent.*

[10] Cf. Hayek (1942, p. 84): "The world of Science might in fact be described as no more than a set of rules which enables us to trace the connections between different complexes of sense perceptions. But the point is that the attempts to establish such uniform rules which the perceptible phenomena obey have been unsuccessful so long as we accepted as natural units, given entities, such constant complexes of sense qualities as we can simultaneously perceive."

3.2.3. Theory and Evidence

With all these possible interpretations rejected, we are yet to learn what makes science unique, and what Hayek was getting at by speaking of the reclassification of "the very sense qualities which most of us are inclined to regard as the ultimate reality". Soon after C) we find the following:

D) "While at first the new elements into which the physical world was 'analyzed' were still endowed with 'qualities', that is conceived as in principle visible or touchable, neither the electrons nor waves, neither the atomic structure nor electromagnetic fields can be adequately represented by mechanical models. The new world which man thus creates in his mind, and which consists entirely of entities which cannot be perceived by our senses, is yet in a definite way related to the world of our senses" p. 84.

Maybe, after all, what Hayek has been claiming all along is that there is no *reduction* of the entities and classes of entities that we decide to posit in our explanations of the world, and of the truth-conditions of sentences that we form using designators referring to those entities, into specific sensory experiences[11]. We posit electrons, waves, fields, incomes, utilities, etc. as theoretical constructs which are put into relations that, *together*, allow us to derive sets[12] of sentences whose truth can be gauged with the help of sensorial inputs, say those resulting from an experiment, but there is no sensorial definition or translation of any of the terms referring to those entities, their relations, or of the truth conditions of sentences one by one: we cannot see, touch, smell or listen to electrons, nor can we see, touch, smell, or listen to the force of gravity that is pushing us to the center of the Earth, nor, usually, can we see, touch, smell or listen to anything that shows that one sentence, *without many ancillary other sentences*, is true. Still, by positing electrons, and all the other things we choose to posit for them to do any work,

[11] Cf. Hayek (1942, p. 84): "Although the theories of physical science at the state which has now been reached can no longer be stated in terms of sense qualities, their significance is due to the fact that we possess rules, a "key", which enables us to translate them into statements about perceptible phenomena."
[12] Hayek does not draw this conclusion regarding sentences. He seems to only go so far as to say that terms have empirical content only as parts of sentences. We are going beyond him in saying that sentences have empirical content only as parts of sets of sentences. But actually, it is only as part of whole languages. These insights are influenced by Quine (1951).

we can explain *to our satisfaction* why, when we flick the switch, light bulbs give out light, or, by appealing to the law of gravity with its accompanying ontology, why when we jump we fall back to the floor.

But the conclusion is the same: nothing here distinguishes science from other forms of inquiry. Electrons are not observable in *any* sense, but neither is the Wrath of God that some have used to explain meteorological catastrophes, or even the peculiar character of the German, or the English, or the French, so frequently used to explain seemingly deep-rooted cultural differences. In fact, if there is something that humans know how to do is the positing of unobservable entities and more or less mysterious conceptual interconnections, causal and other, to explain whatever they believe needs explaining. As Quine (1951, p. 45) says, "science is a continuation of common sense, and it continues the common-sense expedient of swelling ontology to simplify theory."

3.3. Summary

It is time to take stock. Hayek views inquiry as a process of classification and reclassification, with positing of entities, and definition of groupings of entities and inferential relations between classes of entities that are relevant to explaining whatever it is we find perplexing. Inquiry may be more or less explicit, more or less sophisticated, or more or less self-aware. The audience may be *orbi*, as in science, or *urbi*, as in Jane's explanation of car scratches. Still, the gauge of success will always be the judgment of those participating actively and passively in its process. Sometimes, new theories are not only formulated in different vocabularies, but are built with different rules, both inferential and normative, than the ones previously held. The encomium "scientific" is granted to theories when we wish to say that they are respectable, credible, institutionally weighty, and worthy of being publicized across the community. Any demarcation is sociological, not epistemological.

Yet, Hayek thought he needed the distinction between science and something else. The reason is that one of his goals in the *Scientism* essay is to defend that methods should be adjusted to the objects of inquiry. I will show in the next chapter that in order to do this,

he contrasts the methods of the natural sciences with those of the moral sciences by saying that, whereas the natural sciences try to reclassify things independently of their effects on humans, the social sciences will have to take into consideration the way agents classify them. To this part of his essay we now turn.

4. The Moral Sciences

According to Hayek, the moral sciences are "concerned with man's conscious or reflected action" (pp. 88-9). He clarifies that not all the sciences that have a social or a human object of study are moral sciences. There are what he calls "natural social sciences" such as certain branches of epidemiology or neurology. These could be studied with the methods of the natural sciences (p. 88). But with respect to the moral sciences, "the situation is essentially different." (p. 89)

4.1. Vagaries with the Notion of Action

Before we can see Hayek's reasons for thinking so, I have to elaborate on what the proper domain of the moral sciences is. When clarifying what "man's conscious or reflected action" is, Hayek says that these are actions "where a person can be said to choose between various courses open to him" (p. 89). He does not offer any argument, but the relationship between the two formulations is not obvious. I will offer a reading of his ideas that hopefully makes them clearer.

The reason why the relationship between the two formulations is not obvious is that it is possible that an agent *chooses* between various possible courses without reflecting upon or even being conscious of it. An example suffices to establish this possibility. Suppose John proposed to Mary. He did it because he wanted to spend the rest of his life with her and was certain that she would accept. Unfortunately, unbeknownst to John, Jane was in love with him too. When a few months later Jane learned that John and Mary were to get married, her heart was broken. Clearly, John broke Jane's heart, but he did not do it either consciously or reflectively. Still, it is certainly something he *did* by proposing to Mary, which in turn is something he chose to do.

An exploration of this example will help us see how Hayek's two formulations are related. Let us look at the following terms:

a) John's proposal to Mary.
b) John's kneeling.
c) John's producing a ring.
d) John's hurting of Jane's feelings.
e) The hurting of Jane's feelings.

Notice that each of these refers to events happening at a time and place, with specific durations. We might say that when John kneeled, his proposal started, and that it ended with Mary's acceptance. It is also safe to say that meanwhile John produced a ring, and that two months after their engagement Jane's feelings hurt. How many things did John do? How many did he choose to do? How many did he do consciously or reflectively? When and where?

One convenient way to deal with these questions is to say that an action is an event that is in a certain relation to an agent, and that, as usual, as there are many possible descriptions of this or that event, there are many possible descriptions of this or that action. According to Davidson (1971, p. 46), a condition for an event to be someone's action is for there to be *some* description of the event that indicates that the person did it intentionally[1]. When John kneeled in front of Mary in b), it is true to say that he did it intentionally because we can easily give the reason why he did it[2]: he wanted to propose and thought that the ritual of proposal involved kneeling. So the movements of his limbs, which we may describe as kneeling, were an action of John's. He did many other things for the same reason: he looked into Mary's eyes; he asked her if she wanted to marry him; he produced a ring, etc.

[1] Davidson (1971, p. 46) formulates it thus: "A man is the agent of an act if what he does can be described under an aspect that makes it intentional."
[2] The relation between intentionality and reasonability has caveats which I will disregard. Cf. Davidson (1973a, p. 79): "If the agent does x intentionally, then his doing x is caused by his attitudes that rationalize x. But since there may be wayward causal chains, we cannot say that if attitudes that would rationalize x cause an agent to do x, then he does x intentionally." The wayward causal chains to which Davidson refers are situations where our reasons cause us to act, but not in the right way: they do not cause the action insofar as they are reasons for that action, but in some other way. The most famous example of this, the climber's, is Davidson's (1973a, p. 79): "A climber might want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening his hold on the rope he could rid himself of the weight and danger. This belief and want might so unnerve him as to cause him to loosen his hold, and yet it might be the case that he never *chose* to loosen his hold, nor did he do it intentionally."

I have also noticed above that we might wish to say that his proposal to Mary really *is* all these things together, and that it ends with her acceptance, marking the engagement. So, with respect to sentence a), we may say that his proposal to Mary is really one way of grouping all these things that he intentionally did with an eye to becoming engaged.

The two sentences "John proposed to Mary" and "John tried to become engaged to Mary" are different ways of talking about the same actions of John's, one of the results of which is described by "John's engagement to Mary". The difference between the two is that we might feel more inclined to use the one rather than the other in specific circumstances. For instance, we might prefer to use the second if he had failed to become engaged to Mary since, in most circumstances, it implicates that he failed[3].

But there are other consequences to his actions, consequences that John did not intend. His engagement caused Jane's feelings to hurt[4]. So we may also describe John's actions as in d). What is crucial is that d) does *not* refer to the same event as e). Whereas d) refers to *his* hurting Jane, e) refers to Jane's feelings hurting, which occurred two months after his proposal. When did he hurt Jane? If by that we mean "When did he do whatever he did that hurt Jane?" the answer is obvious: when he proposed to Mary. It just happens that Jane felt hurt only two months later. So John hurt Jane the day he proposed, Jane's feelings were hurt two months later, and the event involved in the former caused that referred to by the latter. So d) refers to his proposal by describing it through one of its effects, the same way that something may be described as a poison after its effects on an organism.

The upshot of this discussion is that "actions that are conscious or reflected" does not define a class of actions[5]. We may truthfully say that John's proposal to Mary was conscious or reflected, but we cannot truthfully say that John's hurting of Jane's feelings was conscious or reflected, *even though the two underlined descriptions refer to the same actions*. An action may be conscious or reflected if described in one way,

[3] Cf. Davidson (1967a, p. 101): "Redescription may supply the motive ('I was getting my revenge'), place the action in the context of a rule ('I am castling'), give the outcome ('I killed him'), or provide evaluation ('I did the right thing')."
[4] Cf. Davidson (1971, p. 53): "An agent causes what his actions cause."
[5] Cf. Davidson (1971, pp. 46-7): "we can without confusion speak of the class of events that are actions, which we cannot do with intentional actions."

but not be so if described in another. Further, it is also not the case that when you fill the blank of "_____ is an action that is conscious or reflected" and of "_____ is a choice between various courses" with the *same* description of an event you always get the same truth value in both resulting sentences. Take an experienced driver who took his usual path home and could hardly remember the ride. If asked, he would rightly tell us that he *chose* to take one route rather than another possible alternative. So if we fill the blanks with "the driver's going down 46th street", the first sentence is false, but the second true.

As mentioned, actions may truly be described as reflected, conscious, intentional or unintentional, depending on how we are choosing to single them out. And how we should so choose depends on our goals. Davidson (1971, p. 58) urges us not to confuse "between a feature of the description of an event and a feature of the event itself." For all these reasons, I urge us to read Hayek carefully, perhaps as saying that the moral scientist is concerned with action *simpliciter*. It is also a first step towards reconciling Hayek's stress on action, which, as we have mentioned, involves the notion of intention, with his well-known emphasis on unintended consequences.

4.2. Do the moral sciences explain action?

We have seen that for Hayek the moral sciences are *concerned* with action. But there are passages where Hayek explicitly says that the moral sciences do not, or could not, *explain* action[6].

For instance, Hayek (1942, p. 103 italics in the original) writes that "it is a mistake (…) to believe that their [the moral sciences'] aim is to *explain* conscious action. This, if it can be done at all, is a different task, the task of psychology."[7]In Hayek (1943a, p. 67

[6] In other places, e.g. Hayek (1942, pp. 88-9), he does refer to explanation of action.
[7] Similarly, in Hayek (1943a, p. 67 italics in the original) he writes: "the misunderstanding is that the social sciences aim at *explaining* individual behavior and particularly that the elaborate process of classification which we use either is, or serves, such an explanation. (…) If conscious action can be "explained," this is a task for psychology but not for economics or linguistics, jurisprudence or any other social science."

italics supplied) he clarifies that "what we do is merely to *classify* types of individual behavior which we can *understand*, to develop their classification – in short, to provide an orderly arrangement of the material which we have to use in our *further* task." This further task is the explanation of the "sort of order [which] arises as a result of individual action but without being designed by any individual."[8] Hayek (1942, p. 103)

Hayek (1942, p. 104) offers a famous example of such an order, the spontaneous development of a path through wilderness. Each person trying to get across wishes to follow a route that is safe, fast, and not too tiring. Who wouldn't? Facing virgin bush, the pioneers might have had to think through almost each step, making decisions as they went along. Their behavior left its traces: obstacles removed, foliage cut, stones judiciously placed, and stepped plants offering the comfort of prior human presence. The people coming afterwards may have found these traces more or less salient. As they made their own decisions, it is possible, if not likely, that they seized, knowingly or unknowingly, the improvements of the pioneers, adding to these their own traces. As time went by, all these traces precipitated into a clear *path* which any walker traversing the wilderness will, more likely than not, follow. No one thought *the* path out. The pioneers certainly did not think it globally: they went step by step. They might not even have followed this *exact* route: its present shape may have altered with time. The path is the result of human action but not of design. It is one unintended consequence of people traversing the wilderness.

The previous paragraph encourages the idea that in order to account for the formation of the path the moral scientist does not need to know the particularities of the traversers' psychology, and hence does not *explain,* in this sense, their behavior. What reasons A or B had for crossing the path, or what was salient to their perception, or what inferential tendencies they had is, *in detail*, irrelevant. All we need to know is that there *were* people who *wanted* to cross, *and* that they *wished* to do so in an efficient manner, *and* that they had similar judgments as to what were *reasonable* decisions to take. This, I suppose, comes from "our general knowledge of how we and other people behave in the kind of situation in which the successive people find themselves who have to seek

[8] Cf. Hayek (1942, p. 103): "Insofar as we analyze individual thought in the social sciences the purpose is not to explain that thought but merely to distinguish the possible types of elements with which we shall have to reckon in the construction of different patterns of social relationships."

their way." Hayek (1942, p. 104) "We use the different kinds of individual behavior thus classified as elements from which we construct hypothetical models in an attempt to reproduce the patterns of social relationships which we know in the world around us." (1943 p. 68)

So it seems that the moral scientist does not explain behavior, but categorizes behavior that she can understand, action, into types, and with these categorizations develops theories about social relations. I have argued in the previous section that an action is an event that is in a certain relation to a person, its agent. I have also given a criterion for an event and a person to be related as action and agent: they are so related if there is some description showing that the person did it intentionally. But it seems difficult that the moral scientist could categorize behavior as action of type x without offering some *sort* of description of the agents' intentions. As I show in the next section, this implies that the moral scientist must offer some sort of *explanation* of action after all.

4.3. Reasons as Determinants of Action

The classification of action which Hayek seems to have considered was into types such as walking across a wilderness, or "signaling or hunting, making love to or punishing another person" (1943a, p. 64). Naturally, as he himself notices[9], in order to thus classify concrete actions we need to *understand* what the agents are doing. Hayek writes that

E) "by his actions, *determined by* the views and concepts he possesses, man builds up another world of which the individual becomes a part. And by 'the views and concepts people hold' we do not mean merely their knowledge of external nature. We mean all they know and believe about themselves, about other people, and about the external world, in short everything which *determines* their actions, including science itself. Hayek (1942, p. 87 italics supplied)

[9] Cf. Hayek (1942, p. 93) "A 'word', or a 'sentence', a 'crime' or a 'punishment' is of course not an objective fact in the sense that it can be defined without referring to our knowledge of people's conscious intentions with regard to it." I cannot account for Hayek's use of the word 'conscious' in this sentence.

In this passage, Hayek emphasizes that it is an agent's *epistemic* attitudes - what she knows and believes - that determine her actions[10]. What does 'determine' mean, however?

I believe that two interconnected readings are plausible. A *first* sense of determines is as 'causes': an agent's epistemic attitudes cause her actions. For this reading to work, however, we have to complement what Hayek says: your epistemic attitudes, by themselves, do not cause you to act. To see this, imagine you believe that there is food available in the kitchen. What action of yours does this cause? Well, if you become hungry, then your knowing that there is food in the kitchen, *together with your coming to have a desire for food,* may, with plausibility, be said to have caused your going to the kitchen to get some food.

The causal insufficiency of the epistemic attitudes is noteworthy because it makes clear that we may know everything an agent knows about the external world, and the way she pictures the world, without thereby being capable of formulating any causal explanation of her actions[11]. It is thus not just an agent's epistemic attitudes that are relevant, but her attitudes more generally, insofar as they form reasons for acting.

To be precise, besides the epistemic attitudes, reasons also involve what Davidson (1963, pp. 3-4) calls 'pro attitudes': "Whenever someone does something for a reason (…) he can be characterized as (a) having some sort of pro attitude [desires, wantings, urges, promptings, and a great variety of moral views, aesthetic principles, economic prejudices, social conventions, and public and private goals and values in so far as these can be interpreted as attitudes of an agent directed toward actions of a certain kind], and (b) believing (or knowing, perceiving, noticing, remembering) that his action is of that kind. (…) let me call this pair the primary reason why the agent performed the action."

A *second*, complementary, reading of Hayek's "determines" refers to the fact that what the agent knows and believes, together with her desires, can be said to determine her

[10] He first talks of "views and concepts", but since the relation of these with what people "know and believe" is obscure, I will brush the first formulation aside.
[11] Nevertheless, knowledge of the epistemic position of the agent does allow us to reduce the set of possible explanations. If we know that Jane believes that that there is no chocolate in the kitchen, we know that her going to the kitchen cannot, barring exceptional circumstances, be because she wanted to get chocolate.

actions in the parallel sense of giving it an interesting propositional *content*, and contextualizing action in "the world of which the individual is a part" (E)). Our knowledge of what an agent knows and believes justifies describing that movement of limbs as Jane's going home, or Jane's running an errand, etc., i.e. allows us to classify the agents' actions in the sense alluded to in the previous section.

As Davidson (1963, p. 8) emphasizes, in order to give the reasons for an action we have to portray the agent as a "Rational Animal", to fuse our horizons, and crack into the particularities of the agent's worldview: "Beliefs and desires tell us an agent's reasons for acting only if those attitudes are appropriately related to the action as viewed by the actor. To serve as reasons for an action, beliefs and desires need not be reasonable, but a normative element nevertheless enters, since the action must be reasonable in the light of the beliefs and desires (naturally it may not be reasonable in the light of further considerations)." (Davidson (1978, p. 84))

I would like to remark that rationalization is a staple common-sense explanation, and thus in line with Hayek's appeal to our "general knowledge" of how people behave in this or that situation. As Putnam (1988 [1991], p. 6) put it "no matter how strongly the tides of behaviorism have run, we have never stopped explaining our behavior and that of others in terms of beliefs and desires. We say, 'I went to school today because I knew I had to teach a class,' or, 'I went to the market because I knew we were out of milk, and I wanted milk to put in my coffee.'" In economics these explanations are usually given through some sort of decision theoretic apparatus. We speak of preference relations and probability distributions, and we rationalize behavior by saying that it maximizes some function representing expected utility, i.e. that it is optimal from the agent's standpoint. Not coincidentally, a reading of the classical contributions to the subject reveal a strong zeal to have axiomatizations capture the structure of our intuitive notions[12].

It is thus quite radical to say that the moral scientist does not explain action, *simpliciter*. I believe that what Hayek was trying to say when he explicitly rejected that the moral scientist explains action is that often, as shown in the previous section, we do not need

[12] This is shown in chapter 5.

to be *very* detailed about our explanations. It was not relevant to our account of the formation of the path whether the people were traversing the wilderness to meet their lover, to please their king, or to search for gold. Still, this *is* a fiction of our example. How detailed we need to be depends on the purposes of our research and on the very path we follow through the wilderness of our questions. Suppose we now wanted to explain not just the formation of the path but also how it came to have its precise shape. It may be important to know that, say, people were not just traversing the wilderness, but that they did so penitently in order to quench the Wrath of their Gods. We may now find the hypothesis that people were trying to follow the safest path unpersuasive.

In any event, the fact that the reasons we offer are often not detailed does not mean that they do not *explain* action. Suppose you ask Jane why she did not eat a slice of chocolate cake, and she replies that she did not feel like it. Her answer is trite, but it *explains* her behavior. It is certainly informative. You have learned, for instance, that it was not because she did not *know* that cake was available that she did not have a slice, or because she did not like the way it looked. You may be perfectly satisfied with this answer. Still, if you are inquisitive, you may ask "why not?" She may reply that she is not very fond of chocolate cake. Now you know that next time she visits you should prepare some other dessert. You may go on inquiring: she is not fond of chocolate cake because she is not fond of chocolate, and she is not fond of chocolate because she finds cocoa too bitter, etc. How detailed an explanation you seek depends on your purposes and how well you want to know Jane[13].

Hayek's position on what the moral sciences do raises several questions. First, there are the questions involving what, above, Hayek describes as "the general knowledge of how people behave in this or that situation". What knowledge is this? Also, since to recognize "the kind of situation in which the successive people find themselves", and hence to categorize this or that concrete behavior as action of a specific type we need to

[13] Hayek (1942, p. 103f) gives the example of an economic planner: "while economic theory might be very useful to the director of a completely planned system in helping him to see what he ought to do to achieve his ends, it would not help us to explain his actions – except insofar as he was actually guided by it." This is confused. The very description of him as "director" already attributes intentions, even if not very detailed. I could already say that he did so-and-so, say nationalize the oil industry, because he was planning the economy. He might have nationalized it for altogether different reasons. Maybe he wanted to run naked across the oil fields. In this case he would not have acted as a director (type of action).

rationalize agents' behavior, we need to be told how the moral scientist is to produce the required reasons. To these questions I will dedicate the following sections.

Finally, we have to ask why the moral scientist, if she is interested in studying social relations, is limited to constructing explanations based on this typology of action, and not, say, by directly exploring, in a non-individualistic way, social entities. As Hayek (1943a, p. 69) puts it, "have we not in these social structures at last definite tangible social facts which we ought to observe and measure (…)? Should we not here at least derive all our knowledge by observing and experiencing, instead of by 'constructing models' from the elements found in our own thought?" Indeed, it is not *prima facie* evident why such elements are the descriptions that underlie the most fruitful methods for discovering and explaining social patterns. I will assess Hayek's discussion of these matters in the final section of this chapter.

4.4. We are all Birds of a Feather[14]

I have shown that the moral sciences are concerned with action, and that action is causally explained by agents' reasons for acting. As noticed in 3.3., their reasons and intentions depend on their worldviews, i.e. on the world which the agent makes. The moral scientist thus needs to take into consideration not just what she believes, but also the worldviews, "the working of the mind" (p. 91) of the people she is studying. Indeed, my knowledge that, say, oil results from the fossilization of the remains of zooplankton is incidental to my classifying it as a means of production. It is a means of production because agents have thought out ingenious ways of putting it to all sorts of uses. Similarly, it is not because there *is* oil in North Dakota that oil wells were built there, but because the relevant agents *believed* that there was oil in sufficient abundance to recoup their investment. When I am trying to understand other people's behaviour, I thus need to consider not just how the world is, but how *they* see the world.

[14] This expression comes from Quine (1969b, p. 125) "it is reasonable that our quality space should match our neighbor's, we being birds of a feather."

4.4.1. A Triangular Relation

But how am I to do this? According to Hayek,

> F) In his conscious decisions man classifies *external stimuli* in a way which we know solely from our own subjective experience of this kind of classification. We take it for granted that other men treat various *things* as *alike* or *unlike* just as we do, although no objective test, no knowledge of the relations of these things to other parts of the *external* world justifies this. Our procedure is based on the *experience* that other people as a rule (though not always—for example, not if they are color- blind or mad) classify their *sense impressions* as we do. P. 89 (italics supplied)

Hayek writes that there are things/stimuli that humans group as equal or unequal, but that no knowledge of the *external* world would, of itself, give us reasons for so grouping. We have to exercise care in interpreting Hayek's *external.* After all, the criteria we ordinarily use in order to assess whether two people have similar perceptions are behavioral, and therefore as external to *us,* the observer, as the stimulus is to them: we conclude based on *our* observation of agents' behavior, broadly construed[15], and employing *our* prior standards and theories of *behavioral* similarity that a stimulus is causing similar responses. As Hayek (1943b, p. 109) writes, "it is true, of course, that we know nothing about other people's minds except through sense perceptions, that is, the observation of physical facts."

Notice that we are involved in a triangle: an observer (us), one or more subjects (the agents), and a shared stimulus. I propose to interpret Hayek's 'external' as emphasizing externality to the subject, since the subject's reactions are as external to us, the observer, as the stimulus. Hayek's point would thus be that the observer has, on the one hand, a relation of similarity derived from her observation of a subject's reactions, and, on the other hand, a relation of similarity based on her objectivistic study of the cause of the subject's reactions. These relations, he says, are, in general, different: the subject may

[15] Agents' *verbal* behaviour included. For examples of behavioural criteria of perceptual similarity, see Quine (1973, pp. 16-8).

be found to classify entities in ways that are not isomorphic to the classes that the observer, guided by her study of the stimulus, would construe.

Naturally, the observer is also a perceptual human being. Indeed, Hayek tells us in F) that the observer classifies stimuli the same way the subject does. But then how does this cohere with Hayek's position that there is not, in general, an isomorphism between the two classifications? In other words, what distinguishes a person *qua* observer from the same person *qua* subject, as Hayek's account clearly requires?

The two positions are made coherent by appealing to the goals of inquiry: the way the subject makes sense of the stimulus depends on her worldview, including her purposes, and the same thing should be said about the observer. This is, fundamentally, the same conclusion I arrived at in 3.2.2: that through immediate perception we may not find any relevant difference between, say, two white powders, yet another (chemical or other), differently motivated analysis may lead us to the preferred conclusion, *for those purposes,* that the two powders are different substances In 2), I noticed an undercurrent to Hayek's writing, in that he sometimes downplays the role of the observer, and leaves the subject with her secondary qualities and the object with its primary qualities by themselves: as if the third corner of the triangle were dispensable for the very notion of objectivity[16], as if it were coherent to discuss the properties of an entity outside some (meta)theory[17], as if the observer could rise above all standpoints.

These perplexities lead Hayek to an interesting conclusion: that the observer in the moral sciences is in a particular predicament in that she cannot rely solely on her observation of reality external to the subject in order to *understand* the subject's behavior, because the agent's conceptualizations cannot be recovered from an independent study of the causally relevant reality external to the agent. Yet, the agent's

[16] Cf. Davidson (1995b, p. 18) "All propositional thought, whether positive or skeptical, whether of the inner or of the outer, requires possession of the concept of objective truth, and this concept is accessible only to those creatures that are in communication with others. Knowledge of other minds is thus basic to all thought. But such knowledge requires and assumes knowledge of a shared world of objects in a common time and space. Thus the acquisition of knowledge is not based on a progression from the subjective to the objective; it emerges holistically, and is interpersonal from the start."

[17] Part of this tension is resolved by Hayek's insistence that it is always through *further* observations that we refine our judgments. So Hayek distinguishes what may be called first-order from higher-order observations. The observer and the subject may thus conceivably be the same person, when engaging in different levels of observation. Still, there is a *social,* intersubjective aspect to objectivity that is never taken into account in any of Hayek's works that I am familiar with.

behavior depends on her conceptualizations, on her attitudes. How is the moral scientist to explain behavior? Hayek states that "when we have to explain human behaviour towards *things*; *these* things must (…) not be defined by what we may find out about *them* by the objective methods of science, but in terms of what the person acting thinks about *them*" Hayek (1942, p. 93 italics supplied). By what methods then are we to determine other people's worldviews? To this I now turn.

4.4.2. What Do We Share?

Hayek gives the moral scientist a way out of her predicament: she is also a human being, and therefore not only shares, but *knows* that she shares much with her subjects. In doing moral science, she needs to tap this common nature.

What exactly does she share with her subjects, though? Hayek is unclear about this. In F) he is explicit in saying that the observer classifies *sense impressions* the same way her subjects do. But it is easy to show that this is insufficient for moral scientific purposes. Indeed, there may not be any difference, perceptual or other, between two copies of *The Wealth of Nations* that a *librarian* would find relevant. But one of the copies may be very precious to *Edgar* because it was a gift from a deceased relative. No physical inspection of the two copies, perceptual or more sophisticated, will reveal differences between the two that justify their being placed on different bookshelves: hence our calling them copies. Still, there are descriptions that show the two copies to be different according to other criteria: to Edgar one copy is a memento and the other is not, and that is relevant to explaining why he keeps the first on his nightstand[18].

[18] It is untenable to say that there are strictly no physical differences (i.e., describable in the vocabulary of physics) somewhere that could be compounded to draw a distinction between these two copies. Still, such physical criteria may be far back in the natural history of each copy, or deep inside Edgar's brain and so forever outside any judgment based on direct observation of the two copies, which seems to be Hayek's major point (cf. 4.5. below). Truth is, we can always find *some* difference, in some vocabulary, between two entities "in relation to other things." Two exact book copies may be in a different relation to Edgar's nightstand, for instance: one is on it, the other is not. The question must not be whether we could, in some vocabulary, describe *some* difference between two entities in relation to other things: the fact that we believe that they are two and not one entity implies this trivially.

Edgar does not find the two copies perceptually more or less similar than the librarian, nor do we have reasons to believe that he would doubt us if told that "all objective tests" fail to tell them apart. The differences in significance are far away from Edgar's or our stream of experience. Thus, that the moral scientist classifies sense impressions the same way her subjects do, even if true, is not, *pending further argument*, sufficient to solve the moral scientist's predicament.

Instead of speculating what such further argument might be, let us try to further assess what it is that Hayek believes that we share. Take the exceptions Hayek mentions in F): color-blindness and madness. The first exception is indeed an exception to our classifying sense impressions the same way, but it is a curious exception in that it seems to make absolutely no difference to the moral sciences. I do not think that the behavior of color-blind people is harder to understand than that of someone with 'normal' color-vision[19]. Hence, shared classifications of sense *impressions* seems, again pending further argument, not to be *necessary* for moral scientific purposes.

Indeed, color-blind people have other mechanisms for aligning their use of color words with the conventional usage in their community, and this is all that matters: if differences in the classification of color are behaviorally inscrutable[20], then the differences are irrelevant for most, if not all, moral scientific purposes. As Hayek (1952, p. 31) writes, "nothing can become a problem about sensory qualities which cannot in principle also be described in words." In the *Scientism* essay (1942, p. 86f italics supplied) he writes: "that different people classify external stimuli in the 'same' way does not mean that individual sense qualities are the same for different people (*which would be a meaningless statement*), but that the systems of sense qualities of different people have a common structure."[21]

---

[19] Incidentally, colour blindness is a broad denomination for several possible variations in human colour vision. In many cases the person may be unaware of, usually his, 'impairment' until a colour-blindness test is performed. Indeed such a situation is interesting because the person is only found to have a deficient colour vision when confronted with an admittedly behavioural but still artificial test that enlarges the evidential basis.

[20] This form of behaviourism is *not* the behaviourism of a reductionist type that Hayek (1952) fights.

[21] I am, obviously, interpreting "common structure" as having a behavioural expression. Cf. also Hayek (1952, p. 31): "Most people will agree that the question of whether the sensory qualities which one person experiences are exactly the same as those which another person experiences is, in the absolute sense in which it is sometimes asked, an unanswerable and strictly meaningless question. All we can ever discuss is whether for different persons different sensory qualities differ in the same way. To establish whether a

The second exception he mentions is madness. Yet, except perhaps for hallucinations, 'madness' is not usually related to eccentric classifications of sense *impressions*, but to the propositional contents of a patient's attitudes. A psychotic who thinks she is the Messiah may have a conventional classification of sense impressions. Now, it might be hard, if not impossible, to give a coherent account of the mad person's reasons for doing this or that. But *that* is why we say that someone is mad to begin with! We do not say that they are mad and, *therefore,* that we cannot understand what they do without appealing to the vocabulary of mental illness[22]. In other words, the very decision whether to consider a fellow human being as mad requires that the predicament of the moral scientist be overcome. In fact, this is what makes the social context of madness such an interesting topic for moral scientific explorations.

Naturally, these may be considered illustrative examples with little to no substantive intent. But then what is it that we share? Take the following passages:

G) 1. But what are the consequences of the fact that people perceive the world and each other through sensations *and concepts* which are organized in *a mental structure* common to all of them? Hayek (1942, p. 87 italics supplied)

2. *the knowledge and beliefs* of different people, while possessing that common *structure* which makes communication possible, will yet be different and often conflicting in many respects. Hayek (1942, p. 92 italics supplied)

3. "we must start from what men think and mean to do: from the fact that the individuals which compose society are guided in their actions by a classification of things or events according to a system of sense qualities and of concepts which has a common structure *and which we know* because we, too, are men." Hayek (1942, p. 97)

person is color blind we have to find out, not how 'red' looks to him in any absolute sense, but whether and how it differs from various other shades of 'red' and from 'green'. In all such instances we can find out and know only whether, compared with other people, a person discriminates between given stimuli in the same or in a different manner."

[22] Cf. Hayek (1943a, p. 64): "on watching a few movements or hearing a few words of a man, we decide that he is sane and not a lunatic."

> 4. From *the mental categories* we have in common with them [other human beings] we can reconstruct the social complexes which are our concern. Hayek (1943b, p. 114 italics supplied).

In these quotes Hayek gives many suggestions besides that of our sharing sense impressions. According to G.1. and G.2. it is clear that we share a *structure.* In the former we share a structure of sensations and concepts, in the latter it is our knowledge and beliefs that are similarly structur*ed.* In G.3. we are told that we actually *know* such system. Finally, in G.4., we are told that not only do we have a structure in common, but that we actually share mental *categories.*

It is seldom clear what Hayek means by *any* of these, much less how they are interrelated. He mostly gives examples and expects us to gauge the relevant similarities. For instance, he (1943a, p. 64 italics supplied) writes that "I shall, from a few observations, be able rapidly to conclude that a man is signaling or hunting, making love to or punishing another person"; and that "if I see for the first time a big boulder or an avalanche coming down the side of a mountain toward a man and see him run for his life, I know the meaning of this action because I know *what I would or might have done in similar circumstances*." It is thus, perhaps, not because we *actually* share anything, but by interpreting "on the analogy of our own mind" (Hayek (1943b, p. 139)) that we make sense of other people. In Hayek (1943a, p. 64 italics supplied) we are told that "we thus always supplement what we actually see of another person's action by *projecting* into that person a system of classification of objects which we know, not from observing other people, but because it is in terms of these classes *that we think ourselves*."[23]

[23] Hayek's defense of the vague assumption that we share quite a lot with our fellow human beings is revealing of his overall position. It is found in the following passages:

1. If we are not more aware that (…) we necessarily rely on our own knowledge of the working of a human mind, this is so mainly because of the impossibility of *conceiving* of an observer who does not possess a human mind and interprets what he sees in terms of the working of his own mind. Hayek (1942, p. 91 italics supplied)

2. [The moral sciences] deal with phenomena which can be understood only because the object of our study has a mind of a structure similar to our own. That this is so is no less an *empirical* fact than our knowledge of the external world. Once we have learned that our senses make things appear to us alike or different which prove to be alike or different in none of their relations between themselves, (…) the fact

I think there is a lot of truth to Hayek's account. I do think that humans are all birds of a feather. As illustrated in chapter 2, they seem capable of gearing their perceptions and inductive habits to what makes sense in their several communities. They are also capable, if given time, of understanding each other well. I may not, at first, understand why people with a different culture perform a certain ritual. But I am confident that if I interact with them long enough I will be able to make perfectly good sense of their behavior. I am confident in this precisely because I see them as *like me*, in some hard to spell-out way. I do not think it is because we share mental *categories*, though. I *may*, of course, employ my categories to accommodate their behavior in a theory that *I* find plausible. For instance, I may classify a certain ritual as "punishment", even if I am

---

that men classify external stimuli in a particular way *becomes* a significant fact of experience which *must be* the starting point in any discussion of human behavior. Hayek (1942, p. 92 italics supplied)

3. "But it also follows that it is not only impossible to recognize, but *meaningless* to speak of, a mind different from our own. What we *mean* when we speak of another mind is that we can connect what we observe because the things we observe fit into the way of our own thinking. But where this possibility of interpreting in terms of analogies from our own mind ceases, where we can no longer "understand" there is no *sense* in speaking of mind at all; there are then *only* physical facts which we *can* group and classify solely according to the physical properties which we observe." Hayek (1943a, p. 66 italics supplied)

On the one hand, Hayek says that it is inconceivable that there should be a *mind* radically different from our own. On the other hand, he tells us that our coming to realize that things have properties that are not in a simple relation to those that we attribute to them after their immediate effect on our sensorial apparatus, makes the fact that people classify things in similar ways something we have learned from experience. Whether he is right or wrong, I do not think that Hayek is in contradiction here, for two reasons.

First, we could say that even if it is inconceivable that two minds do not, in general, have similar structures, it could still be an empirical matter to what extent similarity is not equality, how much they need to share for interaction to run efficiently, etc. The same way that we could not conceive that there be no youngest student in a non-empty classroom, which student is the youngest, how older the other students are, etc. are still empirical matters.

Second, Quine (1951) has long taught us that whether a truth is analytic or synthetic is a matter of degree. There are statements in our web of beliefs that we are more reluctant to abandon than others: we may redefine notions in order to keep them true in the face of new evidence, as we choose to throw other beliefs by the board. What used to be 'constitutive' may become only synthetic: it may have been constitutive of our notion of 'mind' that whatever is a mind must share the same basic structure, but as we learnt, so Hayek tells us, that actually the world is not like our senses tell us it is, we redefined mind in a way that their similarity is now an empirical statement about *human* minds.

There is no reason why this diachronic process of theoretical readjustment to new evidence could not be understood synchronically, too: we treat other people's minds as similar to our own as almost a matter of definition on our daily business with them, but as, say, economists worrying about the details of intersubjective interaction, we may treat such statements of similarity closer to synthetic truths. Yet, and anticipating the theme of 4.7., we still want to retain homonymy of expression because our everyday notion of mind and that of the economist are inextricably linked.

So Hayek is not in contradiction, and it is coherent with his position to say that science, as any social endeavor, is impossible without people sharing much, but science can, and has, turned upon itself to explain why and to what extent this is so.

aware that they do not use such concepts. Similarly, as an economist, I will also think of their behavior through the lens of cost-benefit analysis. I also do not think, for reasons expounded above, that it is because we classify impressions the *same* way that we can make sense of each other, although communication would be difficult if we could not *adjust* our inductive tendencies.

We talk to people, we observe what they choose to do in this or that situation, and we try to integrate all this evidence into a coherent theory of this or that person. We never start from scratch, of course: on first meeting a person we select a few promising general theories of who they are based on their appearance, accent, the context of our interaction, our mood, etc. These theories come from our previous experience as social beings, and our personality. With further interaction we adjust the broad theories that were unconsciously selected and, perhaps, if the person plays a frequent part in our life's play, they turn into a custom-made theory making sense of this person alone. As we become ever more confident that we know someone, we feel that we can understand what they are doing, their motivations, their desires, fears, hopes, beliefs, etc., and we develop expectations of their behavior which, in turn, will influence our own decisions.

In line with an understanding of science as differing from every day inquiry as a matter of *degree*, moral science in understanding action does something similar to our everyday interpersonal understanding. The difference is not in kind but in the purposes, and in how articulate and self-conscious moral scientific explanations are. Not that we are not, in our daily life, sometimes conscious of what we think of people and why, or that we do not develop second, and higher-order theories about them. We have to, and how we do this is, without a doubt, a crucial topic in the moral sciences.

Nothing I am saying implies that Hayek is wrong. Quite the opposite. The fact that observation of behavior, broadly construed, in conspicuous circumstances is everything we have to go on does not mean that we could *reduce* our explanations of action to behavioral protocols. In particular, that we could explain action without many analytical hypotheses about what we share with our fellow human beings.

Still, Hayek does *not* offer any constructive account of how we, both as ordinary people and moral scientists, turn a behavioral evidential basis into theories about people. I will try to contribute to this goal in chapter 5.

Before I get to that, however, I will take the opportunity to discuss two features of moral scientific explanation that in Hayek often come across as not present in the natural sciences, and that are of particular importance to economics and to my reflections in future chapters. In the next section (4.5.), I discuss Hayek's claim that the objects of human action are defined by reference to agents' intentions, and not to their structural properties. This will prove useful when I discuss the socialist calculation debate in chapter 6. In the section after that (4.6.), I show that the importance of attitudes to the moral sciences implies that traditional logic, the logic of mathematics, is of limited use in economics. This will prove insightful in the next chapter. I then end this chapter (4.7.) by fulfilling the promise made at the end of 4.3. of assessing Hayek's emphasis on individualistic explanations of social phenomena.

4.5. Structural properties and functional notions

In the previous section we saw, in quote F), that Hayek argues that it is not because things have specific properties, or are in a specific relation to the external world, that agents treat them in one way or another. Pursuing this point, Hayek notices that important moral scientific terms "are abstractions from *all* the physical attributes of the things in question and their definitions must run entirely in terms of mental attitudes of men towards the things" (p. 91 italics in the original). In this section, I will assess to what extent this abstraction from physical attributes is peculiar to the moral sciences, concluding that it is not peculiar at all. I will then derive ontological implications on what to count as an *economic* event from Hayek's thesis that definitions must run in terms of mental attitudes.

Hayek states that "we can choose almost any object of human action" (p. 89) as an example of his thesis. One he explicitly gives is that of a tool (p. 89): something is not a tool because it is made of a specific material or because it has a certain shape.

According to Hayek, something is a tool due to "the use for which it is designed by someone" (p. 90). He urges that "a definition which is to comprise all instances of the[se] class[es] will not contain any reference to [their] substance, or shape, or other physical attribute" (p. 90)

Hayek is telling us that physical or *structural* properties of things are neither necessary nor sufficient for them to be correctly classified as, say, a tool. This does *not* mean that the structural properties of particular things are irrelevant for moral scientific purposes: Edgar may be using this object as a tool *because* it is very sturdy. The point is rather that structural properties do not enter into the definition of the *class* or type of object of human action. Indeed, that this object is sturdy is incidental to its being a tool. Rather, it *is* a tool because Edgar is *using it as* a tool, whatever structural properties this object has.

However, it is easy to show that the necessity of abstracting from the structural properties of things is not distinctive of the moral sciences. Many, if not all, natural sciences have functional notions, i.e., types that cannot be defined by the specific structural properties of their tokens, but by some sort of role, or *function* in a system or order. Take sunburns: a sunburn is definable as a burn caused by exposure to the sun. It is perfectly conceivable that we should have two burns that are identical down to the atom, yet one be a sunburn and the other not. Yet 'sunburn' is a relevant notion for medical science: sunburns are associated with skin cancer, and we know how to prevent them[24]. Besides sunburns, defined by reference to their cause, i.e., to something external to the object - or, say, 'anxiolytic' or 'antipsychotic', defined by reference to their effects - we could give the example of relative terms: the largest mammal in Africa, or the first hurricane of the year. Indeed, we could have a list of an entity's structural properties without learning whether it was the largest mammal or the first hurricane. Yet, both notions are relevant for natural scientific questions. What is distinctive about the moral sciences is that the function or role is related to agents' attitudes.

Notice, however, that when we speak of structural properties, we speak of the properties of some*thing* which has already been individuated. We classify *this* object as a tool, or

[24] The example of the sunburn, used in a different context, is Davidson's (1987).

this lesion as a sunburn. When we speak of functional notions, there seems to be a prior individuation in some *other* vocabulary.

In the case of the moral scientist, we can always ask whether this prior individuation is the moral scientist's or the agent's. Hayek, in the quote I offered in the first paragraph, writes that it is the agent's attitudes towards "*the things"* that are crucial, in which case it seems that it is the agent's individuation that matters. But it seems to me that I could correctly say that someone is using, for instance, a hammer as a means of production without either assuming that the person individuates the hammer, or that the person has the concept of means of production. It may simply be the best way *I* find for accounting for the agent's behavior**.**

There are many interesting issues resulting from the relativity of moral scientific notions to agents' attitudes. For instance is *this* hammer a means of production? From what I have said in the previous paragraph, the hammer is a means of production if it has a certain role in someone's plan. But what if it had that role this morning in Edgar's plans, but no longer has a role of that kind in anyone's plan? Has it *changed* from being a means of production to being economically nonexistent, or some other thing? Yet nothing seems to have happened to the hammer!

I believe we must learn to deal with the paradoxical tone of these issues. Indeed, no change in *the structural properties* of the hammer need have taken place. Yet there has been a change in the economically relevant properties of the hammer. Events are changes requiring explanation. There is thus an *economic* event that requires an *economic* explanation. Hayek (1943b, p. 132) seems to hint at the main point when he writes that "according to the question we ask the same spatio-temporal situation may contain any number of different objects of study."

Before finishing this section, I wish to draw attention to an important epistemological point. It may seem that, since the structural properties of some entity are neither necessary nor sufficient to determine its role in action, we could not find out what such role is by inspecting or studying the concrete entity. But the fact that there are no necessary or sufficient structural properties need not mean that there are no *typical* properties that offer reasonable grounds for determining the role of a particular entity. It

is indeed conceivable that there be a sunburn indistinguishable in its structural properties from another burn, but it is highly unlikely that a dermatologist will *ever* have any difficulty in distinguishing a sunburn. Similarly, if we may never find out that Alice sometimes uses her favorite hairbrush as a means of production at her mother's hair salon *by inspecting her hairbrush*, we know that oil is, usually, a primary factor, or that pasteurization conducted according to the legal standards is a technique of production often employed. After all, in general, and for the big stuff, we seem to know our way around. Hayek (1943a, pp. 65-66) puts it just right when he writes that "as long as I move among my own kind of people, it is probably the physical properties of a bank note or a revolver from which I conclude that they are money or a weapon to the person holding them." We may conclude incorrectly, of course. But that goes without saying.

4.6. The logic of agents' attitudes

In the *Scientism* essay, Hayek had an intuition: that certain statements are true irrespective of the attitudes of people, yet the truth of other statements is not. Whereas the sentence 'The Summer solstice happens in June' is true if and only if the Summer solstice happens in June, the sentence 'Jane believes that the Summer solstice happens in February' may be true even though the sentence 'the Summer solstice happens in February' is false. As he (1942, pp. 91-2) notices, statements like the second are as independent of *the scientist's* attitudes as the first. But their truth is not independent of *everyone's* attitudes. The moral sciences, insofar as they are concerned with action, with the world which the agent makes, have to deal with statements involving attitudes. In this section I study some of the implications this has on the use of formal systems in economics.

This kind of statement that populates moral scientific explanation has interesting logical features. These are not new to us: we have already encountered them in 4.1.. There, we saw that even though "John's hurting of Jane's feelings" and "John's proposal to Mary" both refer to the same actions of John's, the sentence resulting from the replacement of

x in "x was intentional" by the first designator is false, whilst the sentence resulting from the replacement by the second is true. This example shows that there are formulas whose truth depends not just on that to which designators refer, but on the way we designate them: the features we choose to single out the entities in question.

In order to see why this logical feature is peculiar, take the following three sentences:

(1) Hayek read *The Wealth of Nations*.
(2) *The Wealth of Nations* is the work Adam Smith published in 1776.
(3) Hayek read the work Adam Smith published in 1776.

It is obvious that if (1) and (2) are true, so is (3): we did nothing more than substitute one designator, 'the work Adam Smith published in 1776', for another, '*The Wealth of Nations*', that refers to the same thing; and whatever it is that we read does not change because we name it differently.

But now notice the following:

(1) Hayek believes that he read *The Wealth of Nations*.
(2) *The Wealth of Nations* is the work Adam Smith published in 1776.
(3) Hayek believes that he read the work Adam Smith published in 1776.

Here (1) and (2) could be true without (3) also being true: Hayek may not know or believe that *The Wealth of Nations* was written by Adam Smith, or published in 1776. Hence, the substitution of co-referential terms has *not* preserved truth value. The verb "to believe that" has created an opaque context. Instead, to infer (3), we need as a premise something like:

(2') Hayek knows that *The Wealth of Nations* is the work Adam Smith published in 1776.

This sort of opaqueness, or non-extensionality as it is technically called, is indeed typical in sentences involving the propositional attitudes, i.e. attitudes towards propositions: to believe *that* so-and-so is such-and-such, wish *that*, know *that*, etc.

However, that truth-value does not change when we replace one designator by another that denotes the same thing is axiomatic of classical, or first-order logic, the logic of

mathematics, so often used by economists. If, Hayek has been arguing, attitudes are central to moral scientific explanation in general, and to economics in particular, there are methodological lessons to be learned from the opaqueness of these sentences involving the propositional attitudes. Let us elaborate on this.

When an economist wishes to formalize some chunk of economic theory, she must start by constructing a formal language. She states which symbols belong to the language, and gives rules for concatenating symbols into the terms and formulas of the language. Since formal economics typically employs no more than standard mathematics, the formal language used need be no richer than a first-order language, a language with symbols for variables, relations, functions and logical connectives and quantifiers; with formulas quantifying over variables, but not over relations.

A formal theory is composed of a set of sentences in a specified formal language, and of rules of inference. Its sentences may be divided into axioms and theorems, the latter being the sentences derived from the axioms by the rules of inference. The axioms of a theory can usefully be divided into two types: logical and non-logical. The axiom schemes of the former kind are those that are shared by all theories with the same underlying logic, whereas the latter are additional, and specific to the concrete theory.

One type of first-order logical axiom, the logic of standard mathematics, is an equality axiom of the form $x_1 = y_1 \rightarrow ... \rightarrow x_n = y_n \rightarrow px_1...x_n \rightarrow py_1...y_n$[25], where the *x* and *y* are any terms and *p* any *n*-place relation. A translation of an instance could be "If Jane is the first woman who stepped on Mars, then if Jane is blonde, the first woman who stepped on Mars is blonde, too". It is clear that this *logical* axiom was violated in the example with Hayek and *The Wealth of Nations*, above.

In the case of economic theories, the non-logical axioms reproduce the constitutive syntactical properties and interrelations that economic notions have. For instance, it is typically assumed in economics that preferences are transitive.[26]. When the economist

[25] This formulation of the axiom scheme is found in Shoenfield (1967, p. 21).

[26] Many notions are like this. For instance, a person who uses 'heavier than' to stand for a relation that is not transitive might not be using it to stand for our intuitive "heavier than" relation, but for something else. This merely means that, if we put Mary on one arm of the scales and Jane in another, and observed that Mary was heavier than Jane; we did the same to Jane and Joan, and observed that Jane was heavier than Joan; and then to Mary and Joan and observed that Joan was heavier than Mary; most of us would

uses mathematical theories, she also, usually implicitly, assumes the non-logical axioms that imply the mathematical theory she employs[27].

What I am urging in this section is that it is not just the non-logical portion of our theories that needs to be adjusted for the economists' purposes. *Sometimes it is the very underlying logic that is incapable of reproducing the inferential patterns we wish to preserve*. Above, I identified a counterexample to a first-order logical axiom by using sentences involving the propositional attitudes. I have thus made an, admittedly simple and illustrative, case for finding other logics that better deal with these notions[28].

There are four points I wish to make. In chapter 2 I discussed how the sort of explanations we find relevant depends on the context and on our purposes. I now submit that some inferential relations in economics will always be loose, that the economic is 'open'; which casts doubt on whether formalization of a certain kind is reasonably expectable. Still, there are more ways to be *form*-al than with a first-order language and theory. Other logics have been developed with different syntaxes and semantics that have achieved some success at formalizing our common-sense inferential patterns involving these notions[29].

However, the second point I wish to make is that, although these logics have indeed been developed, and appeared in economics, for instance in game and decision

---

conclude that we had made some mistake when comparing their weights, or that the scales was not functioning properly, as opposed to thinking that we had found a counterexample to the transitivity of 'heavier than'. We would, very reasonably, reject our observations and preserve our theory. In different circumstances we may prefer to simply reject that weight is transitive. This attitude towards sentences is already present in the work of some of the logical positivists. Cf. Neurath (1932, p. 203): "The fate of being discarded may befall even a protocol sentence."

[27] An example is an extensionality axiom like $\forall x(x \in y \leftrightarrow x \in z) \rightarrow y = z$. It says that two sets are equal if their elements are equal. Necessity follows from an equality axiom of the form given in the text, replacing *p* with the mathematician's 'belongs to'. This particular rendition is adapted from Quine (1969d, p. 30). Extensionality axioms like this appear in any axiomatization of set theory.

[28] It need not be *absolutely* impossible to validly formalize sentences involving propositional attitudes with a first-order theory, especially if we are not looking for a general theory but for a stylized theory for a specific purpose. For instance, we could perhaps translate = not as "equals" but as 'Hayek believes to be equal' and hence '*The Wealth of Nations* $\neq$ the work Adam Smith published in 1776' might be true. We could then axiomatize a finite number of Hayek's beliefs, the ones that we wanted to figure in our simple, concrete theory. Human ingenuity knows few limits. See Quine (1977) for a display of such ingenuity. Still, it has usually been found more fruitful to develop other systems of logic that better reproduce the inferential relations between statements involving these recalcitrant notions.

[29] For an overview of epistemic and doxastic logics see Rescher (2005) or Meyer (2003).

theories[30], when economists talk about formal systems, or the importance of 'being precise', or 'doing serious theory', they are usually implicitly referring to standard mathematics, and therefore to first-order theories. Hence, the difficulties here noticed cut deep.

But, and this is the third point, most logics have been developed to try and explicate intuitions of ours. Even if we are not interested in formal theories, the features that these formalizations reveal are often not the result of *the process* of explication or formalization itself, but of the very notions with which they are concerned. This discussion is thus, *pro tanto,* general.

Finally, it is important to note that these logical features are *not* the sole property of the moral sciences. Natural sciences also have to deal with their share of non-extensionality. Take dispositional terms, such as solubility[31]. Something is soluble in water if and only if it *would* dissolve if placed in water. We could try to capture this by a simple material conditional and quantification: something is soluble in water if, and only if, it always dissolves if it is placed in water. This is equivalent[32] to saying that something is *in*soluble if, and only if, sometimes it is put in water, and it does not dissolve. But something may be insoluble even if it *never* has or ever will be placed in water[33]… Other notions that are not easily captured by first-order logic are modalities

---

[30] E.g. Rubinstein (1998, Ch.3).

[31] Cf. Quine (1973, p. 8 italics supplied) "The disposition is a property, *in the object*, by virtue of which the circumstances *c* cause the object to do *a*. The 'by virtue' here is what defies explication. An extensional conditional, a universally quantified material conditional, does not bridge the gap."

[32] $\left[\forall x(A(x) \rightarrow D(x)) \leftrightarrow S\right] \Leftrightarrow \left[\neg\forall x(A(x) \rightarrow D(x)) \leftrightarrow \neg S\right] \Leftrightarrow \left[\exists x(A(x) \wedge \neg D(x)) \leftrightarrow \neg S\right]$, where '*A(x)'* means that the object is put in water at date *x*, '*D(x)'* it dissolves at date *x* and S that the object is soluble.

[33] Quine, a confessed extensionalist, offers, in Quine (1973), an interesting understanding of dispositions: they are notions which we need at present, but that we hope to get rid of in the future. By this he means get rid of them piecemeal: present disposition terms will be explicated away, but certainly others will appear in their stead. Quine believes, in this (1973, p. 10 italics supplied) work, that dispositions refer to properties "in the object": "Where the general dispositional idiom has its use is as follows. By means of it we can refer to a hypothetical state or mechanism that we do not yet understand, or to any of various such states or mechanisms, while merely specifying one of its characteristic effects, such as dissolution upon immersion in water. There are dispositions, such as intelligence, whose physical workings we can scarcely conjecture; the dispositional characterization is all we have to go on. Intelligence is the disposition to learn quickly, if I may oversimplify. By intelligence I still mean some attribute *of the body*, despite our ignorance concerning it; some durable physical state, perhaps a highly disjunctive one. A term for this attribute is entitled to a place in our theoretical vocabulary, even if all we know about the attribute is that an animal that has it is quickly conditioned."

like (some forms of) necessity and possibility: it is necessarily the case that 8 is greater than 7, but are the number of planets in the Solar System necessarily greater than 7? Could not the Solar System just have 5 planets instead of 8? Yet 'the number of planets in the Solar system' and '8' refer to the same thing in the actual world.

The upshot of this discussion is similar to that in the previous section. When Hayek (1942, p. 95) writes, for instance, that "what is relevant in the study of society is not whether these laws of nature are true in any objective sense, but solely whether they are believed and acted upon by the people", it is again the allusion to agents' attitudes that distinguishes the moral sciences, *not* the logical features brought by the opaque context which such attitudes introduce. However, it does seem to be the case that these logical issues are much more serious in the moral sciences than in the natural sciences because of the centrality which sentences with propositional attitudes have in the former.

4.7. Why individualism? Why not?

At the end of 4.3. I left open the question why, as Hayek defends, the moral scientist, who tries to study social relations that transcend any individual's ken, *should* base his explanations on individual behavior, as opposed to studying these social relations or social wholes directly.

It is the 'should' that warrants explanation. No one would quarrel with the thought that individualistic explanations make sense in this or that game, for this or that purpose. Similarly, one might expect that the same could be said about non-individualistic explanations. It is thus clear that, for Hayek's position to be defensible, he has to show either that such explanations are impossible, or that they are uninteresting.

Now, it is undeniable that we have intuitions regarding wholes. We often talk of the state, of society, or of the justice system. It is also undeniable that we often talk as if we attributed agency to such entities. We utter sentences like "the state should tackle inequalities" or "it is society's duty to ensure a minimum level of capabilities to each person" or "the justice system found them not guilty". Still, such sentences are usually a

short way of expressing beliefs that do *not* treat collectivities as agents. For instance, in the case of the examples I gave, it seems that the first could be paraphrased as "the prime-minister should propose measures to tackle inequality"; the second as "we should all fight for other people's entitlements"; and the third as "the judges ruled that they had insufficient evidence to convict them". I believe that most of us would assent to one such paraphrase, and that they are usually easy to derive when the specific game is taken into consideration. Nevertheless, it is undeniable that sometimes we do *not* know how to spell one paraphrase out, and that sometimes such collectivities do take up the role of agents, as Hayek (1943b, p. 119) notices.

Hayek (1943b, p. 123)also notices that if we tried to "distinguish wholes by objective criteria", "if we merely observed the actions of men as we observe an ant heap or a beehive" and "we were really to dispense with our knowledge of what things mean to the acting men", "in the picture such study could produce (…) there could not appear such things as means or tools, commodities or money, crimes or punishments, or words or sentences; it could contain only physical objects defined either in terms of the sense attributes they present to the observer or even in purely relational terms."

Hayek does not deny that social wholes may exist. In fact, he (1943b, p. 119) writes that "the wholes about which we speak exist only if, and to the extent to which, the theory is correct which we have formed about the connection of the parts which they imply." Hayek's position is rather that both the moral scientist and the ordinary person (p. 119-120) "*constitute* these wholes by constructing models from the familiar elements" (italics in the original), and these familiar elements are our interpretations of the roles that concrete things take, and of the meaning of concrete people's behavior.

Clearly, the fact that the wholes that we tend to use are constituted out of our knowledge of individual action does *not* mean that we could not create wholes in other ways. Hayek (1943b, pp. 124-5) writes "that the statistics of words can tell us nothing about the structure of a language will hardly be denied. But although the contrary is sometimes suggested, the same holds no less true of other systematically connected wholes such as, for example, the price system." But why is this true? We may find many patterns by studying the statistics of words, and the same might be true of the price system: just jot down a set of 'prices' from your local grocery store and have a

powerful computer find correlations of some kind with, say, GDP-as-measured-by-the-IMF. You are bound to find something, and I would not be surprised if your findings were quite robust.

Still, whatever it is you find, its interest seems nil. And what Hayek says shows in part why this is so. A procedure for constructing wholes that makes no reference to people's attitudes would not be able to construct the notions that matter. It might define "punishment", or "money", or 'price system', to use Hayek's examples, in *some* way. But this way would not be related to our intuitive notions of punishment or money or price. Unless we associate our theoretical concepts to our everyday, often homonymous, concepts, and especially to our everyday understandings and practices, our theories are simply uninteresting.

For instance, who cares whether capital, a common word with intimations, exploits, another common word with intimations, labor, yet another common word with intimations, unless we can somehow relate this to capital*ists* and labour*ers* as we know them in our concrete conditions of time and place, and to exploitation as something negative, which the victims resent? Who cares about the materialist conception of history, or any other theory, unless it can be related to what happens in the actual venues where our lives are played?

Indeed, Marx himself, who is far from being known for his individualism, spends vast swathes of his *Capital*, especially of the first and only finished volume, illustrating his theses with examples from concrete situations, occurring in specific firms, and referring to actual labourers and capitalists. As Blaug (1997, p. 259) argues, Marx did not mean these illustrations to serve as defenses of his positions, but rather to "build up a graphic picture of capitalist society." They had the powerful rhetorical *and substantial* effect of bringing his reflections down to the concrete practices of people, and by so doing he was able to associate whatever "collectivity" he employed with our common sense individualistic intuitions.

With respect to the ontology of wholes, it seems compatible with Hayek's position to say that wholes need not be *reducible* to individualistic 'foundations'. In other words, that it is not true that any statement about wholes could always, *without remainder,* be

paraphrased with statements implying only the existence of individual action. The argument is rather that there must be some *way* of bringing in the concerns of individuals'. Take the example of the path, given in 4.2.. It seems dubious that we could ever reduce the path to specific actions of specific people. Simply, the specific actions that historically led to the path were in an important sense accidental: we could have changed them in many ways and there would still *be* a path, even if not exactly the same. Still, there must have been individual actions moved by certain general reasons, namely involving a desire to traverse the wilderness. Indeed, Hayek (or anyone else) is *not* interested in *any* unintended consequence of people's actions. They must have some interesting effect on people's lives. This was the case with the path: it resulted from people's wishes to traverse the wilderness, and once it was sufficiently definite, it became a useful route for future traversers. The path is thus an unintended consequence salient enough to merit explanation. There are, of course, relevant consequences that are not salient, but are discovered by the moral scientist's efforts. The role of the price system in economic coordination, or the discovery of routines, customs, conventions, etc. that are crucial, yet often far from conspicuous, are examples of this.

For all these reasons, I take methodological individualism to mean nothing more than the necessity, on pain of irrelevance, of associating the results of our reflections to the practices of individual people. Again *associate*, not *reduce.* Naturally, what strikes us as interesting and relevant depends on our social context. I fully accept that the *interest,* not the validity, of any economic explanation is relative to an audience, and hence to a community. But we may always ask whether the individualism here involved is *itself* a figment of our social context. I do not think so. I believe that the sort of belief-desire explanation explored in 4.3, and the intuition that only *bodies*, viz. human *bodies,* can act are so ingrained in our ways of thinking that the possibility that it might be radically different in other human communities beggars belief.

In any event, my audience and that of Hayek's are certainly individualistic, and thus have no interest in collectivities that are not relatable to the concerns of individuals.

4.8. Summary

In this chapter, I have explored Hayek's views on the moral sciences. I have elucidated what action is, and noticed its relation with the notion of intention. I have also discussed what it might mean to say that the moral sciences study action in its intentional and unintentional dimensions, and I have indicated that what is distinctive about social matters, as opposed to natural matters is the importance of agents' attitudes to the former.

I have also shown that to study action we need to fuse our horizons with those of the agents'. Hayek says that we can do this because we have the fulcrum of our own humanity. Unfortunately he does not spell out in detail how we are to do this. In the next chapter, I will study approaches that try to develop theories about agents by observing their behavior in public circumstances.

5. Revealed Preference and Decision Theory: explicating Hayek's intuition

In the previous chapter I argued, with Hayek, that moral scientific explanation is concerned with action, which, in turn, causally depends on agents' attitudes, i.e., on what they believe and desire. As seen there (cf. 4.4.1.), Hayek would not deny the trivial truth that all the evidence we have for understanding others is behavioral: whatever we learn about other people must come from what we see, listen and read them do in particular circumstances. But I also noticed that, in the works surveyed, Hayek does not offer us an explicit, constructive account of how we use this evidence to build theories about other people. He tells us that we share a lot, that we project our own thought onto others, but he never develops or clarifies these views.

Fortunately, such problems have an old lineage in economics. Revealed preference and decision theories have long tried to base explanation of behavior on intersubjectively available cues, and to behaviorally characterize[1] reasons for acting. In this chapter I go through some of this literature, and explore the evidence we have, and the procedures we do or could follow to turn evidence into knowledge of other people.

I start, in 5.1., with Vilfredo Pareto's discussion of utility functions. Although Pareto treats the problem of determining desires from behavioral evidence as trivial, and therefore does not help us make much *substantive* progress towards our goal, his discussion of the extent to which utility functions are determinate helps me articulate an important ontological point, and it illustrates a methodological position which will be of service in the understanding of revealed preference and decision theories. Since early revealed preference theory tries to account for one form of behavior (consumer behavior) using specific forms of evidence, in 5.2. I survey this literature. Important lessons are drawn, one of which is the peril of insulating kinds of choice behavior, such as consumer behavior, from behavior in general. This difficulty is surmounted by later revealed preference theory. In 5.3. I discuss its canonical formulation, given by Marcel

[1] To use an expression in Richter (1971).

Richter in the late 1960s and early 70s. Richter's framework seems to give part of our fundamental pursuit: he seems to offer a characterization of rational preferences using only observed choice behavior as evidence. Alas, I will conclude that his framework is beset with fundamental difficulties, and argue for the pertinence of introducing verbal reports in order to deal with some of them. The introduction of verbal reports, however, requires a theory for dealing with such reports: we introduce more evidence but we increase the number of variables from two (beliefs and desires) to three (beliefs, desires, and meaning). In 5.4. I study to what extent certain specific syntactical properties are constitutive of economic notions, and I am led to the conclusion that the indeterminacies in our understanding of others imply that we must *choose* a theory from many plausible candidates, each theory with potentially different normative implications.

5.1. Pareto's *Manuale*: an exercise on the ontology of utility functions

I start my survey with Pareto's *Manuale* (Pareto ([1909] 1966)). In this work, Pareto explores the extent to which we can individuate utilities from publicly available evidence (Cf. Pareto ([1909] 1966, first footnote in the Appendix)). In the latter category, Pareto includes verbal reports and the observation of actual choices. It is clear that he believes that this evidential basis is sufficient for us to decide whether *or not* an agent, call him Edgar, is indifferent between any two bundles of goods, or prefers one to the other.

Unfortunately, Pareto does *not* give us a constructive procedure that enables us to derive an agent's preferences from such evidence[2]. Although, for this reason, Pareto will not help us make much *substantive* progress, his discussion serves to introduce ontological and methodological reflections that will prove useful in subsequent sections.

[2] It is also not clear whether he believes, as so many after him, that the observation of choice behavior is alone (without verbal reports) sufficient to recover the preferences and indifferences between *any* pair of goods. See III.52 where he elliptically describes the recovery of indifferences and preferences, and IV.32 where he seems to contemplate verbal reports of introspective judgment as acceptable evidence for making finer distinctions than those offered by actual choices.

One of the ways to represent an agent's preferences, or desires, is by a utility function. Assuming, with Pareto, that we can recover Edgar's preferences and indifferences between any pair of bundles of goods from his verbal and choice behavior, the following mechanical procedure is well-defined:

List the finite set of all possible bundles of goods[3], pick an arbitrary bundle from this set, go through all the bundles in the set, and select those which Edgar is observed to find indifferent to the first. This procedure defines[4] a set of bundles (vectors of goods), call it 1, among whose elements the agent is indifferent[5]. Now take from the original set of all possible bundles of goods the elements in 1, and follow the same procedures, thus defining a set 2. Do this until no element remains. You will end with a set of sets of bundles with empty pairwise intersections, and union equal to the set of all possible bundles[6]. Call these Edgar's sets of indifference.

From each set of indifference *i* choose one arbitrary element, a_i. Now write a sequence thus: pick any two of the a_i, have Edgar compare them, and place the most preferred one to the right of the other. Take another element. Observe whether it is preferred to the rightmost element. If you conclude in the affirmative, place it immediately to its right, if not, compare it to the element immediately to its left, and follow the same procedure until you either find an element to which it is preferred, or there are no elements to the left, in which case place the element there. This finite sequence clearly orders the chosen bundles from least preferred to most preferred[7].

Now let us set up a function U: think of any real number X, and assign it to the leftmost element in the sequence. Then think of any number larger than X, and assign it to the

[3] I am simplifying, since we typically cannot pick up bundles of goods, or put bundles next to each other in a sequence. But, since I am assuming finiteness throughout, I could easily give a name to each possible bundle, and therefore when I say that we are picking a bundle up I should be understood as picking up a name, writing a name next to another in a sequence, etc. In any event, we need not, for now, discriminate between verbal reports and actual choices. So our subject's 'observable' behavior may involve answers to question, and actual choices between pairs of bundles. As mentioned, Pareto seems to accept both types of evidence.

[4] More accurately, "can be used to define". See previous footnote. I will not make further reference to these simplifications in the exposition.

[5] I am assuming that if a bundle *a* is indifferent to a bundle *b*, and a bundle *b* is indifferent to a bundle *c*, then *a* is indifferent to *c*. (transitivity) Also, if *a* is indifferent to *b*, then *b* is indifferent to *a*. (symmetry)

[6] I am assuming a bundle to be indifferent to itself (reflexivity).

[7] Assuming transitivity holds. If it does not, the order in which we pick elements up is not irrelevant.

second element in the sequence. Do this until you have assigned to every element in the sequence a number. Now assign to every element in each *i* the same number which you assigned to a_i. This function is well defined, and has the following properties[8]:

(i) For all x, and all y in the set of possible bundles, U(x)>U(y) if and only if Edgar prefers the bundle x to the bundle y.

(ii) For all x, and all y in the set of possible bundles, U(x)=U(y) if and only if Edgar is indifferent between the bundle x and the bundle y.

We could therefore follow the received nomenclature and baptize U as *a* utility function of Edgar. Not *the* utility function, of course. Looking at the procedure I have sketched, it is obvious that our choice of image set for U was, to a large degree, arbitrary. If instead of X we had chosen any other number, and had followed the procedure from there on, we would have ended up with a different function, call it M, having the same domain, but a different image set, and also satisfying (i) and (ii) with an M where U is found. More generally, given any monotonically increasing function F, from a set of real numbers to another set of real numbers, the function F(U(x)) would also satisfy (i) and (ii) (suitably rewritten, of course), and could be obtained through our procedure by choosing a suitable initial X.

This procedure builds an index that reproduces the order properties of Edgar's preferences and indifferences, assumed to be recoverable from the evidence allowed. The procedure builds an index, not a *measure.* Our knowledge of Edgar's preferences and indifferences *and* our procedure justify us in saying that a bundle *a* brings greater satisfaction to Edgar than a bundle *b* iff U(a)>U(b), for our particular U(x), but they do *not* justify us in saying that bundle *a* gives Edgar U(a)-U(b) more *units* of satisfaction, or even U(a)/U(b) the satisfaction a bundle *b* gives[9].

But could not another procedure give us a measure? For instance if, when we were setting up the function U, we had given the following instructions: "think of any

[8] Assuming the conditions in the previous footnotes.

[9] This is easy to show. Let U(a)=2 and U(b)=1, then U(a)-U(b)=1 and U(a)/U(b)=2. Now let F(x)=x^3. F(U(a))=8 and F(U(b))=1, then F(U(a))-F(U(b))=7 and F(U(a))/F(U(b))=8. But F(U(x)) is as much a result of our evidence and procedure as U(x), and so there is nothing *in our theory* to justify a choice between the two.

positive number X, and pair it with the first element in the sequence. Then pair 2X with the second element in the sequence, nX with the nth, etc." why would this not entitle us to find U(a)/U(b) meaningful as a ratio of satisfactions[10]? After all, only functions obtained from U(x) by transformations of the type F(x)=*a*x with *a*>0 would be turned out by our new procedure, and that ratio is invariant to such transformations.

The problem is that this procedure *seems* arbitrary, and hence we do not feel justified in interpreting the functions it turns out as a measure of satisfaction. All we know about Edgar is a list of bundles such that those written below another are preferred to it, those written to its sides are indifferent to it, and those written above are less preferred than it: there is nothing else besides "above, below, at the same level". So whatever determinateness we get besides this is added by our procedure, not from our knowledge of Edgar's preferences and indifferences between bundles, *and hence* would need further justification.

But can we not make sense of a measure of satisfaction? Do we not have intuitions regarding these? Is it foolish to say that I enjoyed myself twice as much doing *a* than I would have doing *b*? *If* we enlarge the evidence we are willing to accept, or at least explore verbal reports more thoroughly, might we not justify finer distinctions? Pareto thinks so. He says (IV.32) that we can recover not only preferences and indifferences, but also compare *differences* between bundles: we can usually make sense of people's telling us that their lot improved more when they were offered a house when they were homeless, than when they were offered a car when they already had a house. We could therefore update our procedure to deal with this enlarged knowledge about Edgar, and get functions that also satisfy the following:

(iii) For all x, y, and z in the set of possible bundles, U(x)-U(y)>U(z)-U(x) if and only if Edgar prefers to go from having bundle y to having bundle x, than to go from having bundle x to having bundle z.

Now our utility function could undergo positive monotonic transformation of its *differences*. Yet, this is as far as Pareto in the *Manuale* is willing to go. He does not

[10] It should be noticed that this procedure actually happens to make differences and ratios significant since from these two we could easily conclude how many indifferent sets of bundles there are between any two bundles.

believe that there could be *any* evidence, even verbal reports of introspective thoughts, out of which we could justify finer comparisons: for instance to say that we enjoyed the movie we watched yesterday twice as much as we enjoyed the one we watched last week, or even to say that the increase in our satisfaction when we went from homeless to having a house is twice the increase in our satisfaction when we already had a house and went from not having a car to having a car. So it is clear that for Pareto, *pace* some of his predecessors and many of his successors, any *measure* of utility is meaningless. But nothing is lost: he argues that ordinal utility is all we need to do economics[11].

I believe we can draw two important lessons from Pareto. The first is methodological. Notice that our procedure is based on a previously given, intuitive notion of preference and indifference[12]. Pareto believes, better, he takes it as obvious, that we *can* recover people's preferences and indifferences from verbal reports and observation of choice behavior, all the evidence that is available to economic agents in their ordinary intercourse. But this evidence is insufficient to make utility as determinate as other intuitions of ours might have wanted: in particular, we may often talk as if there were a measure of utility, yet such a measure cannot be accounted for when all the evidence is in[13]. Pareto urges us to get rid of such notions.

But there is another, related, lesson, an ontological lesson. If all a utility function does is translate the language of "Joan prefers bingo to chess" into the language of "$U_1(a) > U_1(b)$" where '1' refers to Jane, 'a' to bingo and 'b' to chess, why bother with the much less intuitive, and error-inducing notion of utility at all? One reason I can think of is technical: it is easier to work with utility functions than with preference

[11] Samuelson ([1947] 1983, p. 99) puts the general spirit nicely: "It is as meaningless to argue that one particular utility index is really the true measure of utility as it is to argue that the earth really revolves about the sun and not vice versa. Only in terms of observations other than those envisaged in our market place can a cardinal utility magnitude be defined." In Pareto's case we should, rather, say: "only in terms of observations that give us further information than verbal reports and observed choice behavior can, etc." Hicks (1946 especially the last section of the appendix to chapter 1), as is well known, later argued for the sufficiency of ordinal preferences for economic theory.

[12] The explicit assumption of transitivity, reflexivity and symmetry is really an implication of this more fundamental assumption. I have already explained in section 4.6. what it means to say that a property is constitutive of a concept. I will discuss below to what extent these particular (syntactical) properties are constitutive of our intuitive concept of preference and indifference.

[13] Naturally, we may disagree on how determinate a notion actually is either by accepting more evidence, or by choosing to give primacy to certain intuitions over others. As I showed above, Pareto puts his cards on the table: he tells us point-blank how far he is willing to go before *his* intuitions start wobbling.

relations. Therefore, there is hope that insights will be obtained from working with utility. But whatever the reasons, what the procedure has shown is that we can, in our metatheory, do without any talk of utility: we use utilities in the *theory* and get rid of them in the metatheory by using sentences like (i) and (ii) above to translate them away. As long as utilities are reducible to preferences in this fashion, our theories need *not* be understood as committed to the existence of utilities, only to that of preferences[14].

This short survey of Pareto's work has set the methodological motive for the theories I will explore in the next sections: only accept a notion that can be translated into other notions which you can account for with acceptable sources of evidence, using procedures with desirable properties. Unfortunately, Pareto does not tell us *how* we are to recover an agent's preferences and indifferences from the evidential basis he condones. In the next section, I will discuss early revealed preference theory which tried to explain one form of behavior, consumption behavior, by appealing to well-defined sources of evidence.

5.2. Early Revealed Preference Theories

In its earliest formulations, what is known as Revealed Preference Theory (RPT) was a theory of *consumer* behavior. I will distinguish two versions of RPT[15], a strong and a weak project, elements of which can be found in most relevant contributions from the late 1930s to the early 1950s.

The strongest, and also earliest, project set out to build a theory of consumer behavior without employing the notion of preference, instead using only the notions peculiar to any theory of consumer behavior properly called, namely those of price, income, and good. This was in opposition to the received theory of consumer behavior, derived from the notion of preference, or its index, utility, and belonging to theories of action in general. This early approach, the founding example of which is Samuelson (1938b), was not trying to *reveal* preferences at all, but to get rid of that suspicious notion for the

[14] These ontological reflections are mine, not Pareto's.
[15] That there were at least two formulations is by no means original. See Wong (1978)

purposes of consumer theory. This strong project is interesting to us because, if we could do economics, or just a part of it, without appealing to theories of action, or to the causally related notion of reason, which involves preference, then quite a lot of what I said in chapter 4 would have been irrelevant. I will discuss this project in 5.2.1.

The weak version, on the other hand, had no hopes of *replacing* the old approach with a new one that did not involve any suspicious notion irreducible to behavioral evidence. Instead, it tried to derive sentences, employing only the notions peculiar to a theory of consumer behavior, and that could be used as observation sentences for consumer theory, i.e., that give all the empirical implications of the theory. This will be dealt with in 5.2.2.

5.2.1. The Strong Project

5.2.1.1. Behavioral Problems with the Notion of Preference

Our reflections until now have already shown that action and ancillary notions, such as preference, are beset with difficulties. It is thus easy to understand why the replacement of the older theory of consumer behavior based on preferences with something else may have been found desirable.

Indeed, if by preferences we mean what agents *tell* us they prefer, then the preferences they verbally report need not be causally efficacious, even when the agent holds appropriate beliefs. There are many examples of this: we may honestly report that we prefer one bundle to another, but when facing an actual choice between the two, we realize that we feel differently than we expected, or we may admit that a certain course of action is not as good for us as another, but lack the willpower to do the right thing. This means that the agent may *honestly* report a preference for A over B, and still choose B, even though she believes A to be available too. We may thus say that preferences reported are different from those preferences that are such that the agent chooses A over B (under appropriate circumstances) iff the agent prefers A to B.

On top of this, we have the procedural problems that agents may lie to us, that they may misunderstand our questions, or that we may misunderstand their answers. Together, these difficulties show that the use of verbal reports to recover preferences that have causal relevance to action requires appropriate theories for its successful implementation.

Early Revealed Preference theorists were leery of using verbal reports. However, if we sidestep these difficulties and reduce our evidence to observed choice behavior, other problems arise, viz. whether we could distinguish preferences from indifferences to begin with. When in 5.1. I described a procedure for deriving a utility function for Edgar, I made essential use of Pareto's assumption that we could tell whether *or not* Edgar preferred A to B or was indifferent between them. That "or not" is crucial, marking the difference between decidability and semi-decidability.

In order to see the difference, suppose that we have reasons to believe that *if* Edgar prefers one bundle to another, he will always end up choosing his preferred one, and that we can tell, based on his observable behavior, that he has chosen one over the other. This might give us confidence in adopting the following procedure to recover his preferences: "conclude that Edgar prefers a bundle A to a bundle B if Edgar, when offered both, *eventually* chooses A over B." Since choice is never instantaneous, any actual implementation seems to require something like that "eventually". But this procedure is open to a significant objection. It implies that *if* Edgar does prefer A to B or B to A, then we know (*ex hypothesis*) that he will eventually choose one or the other. But after any amount of time without our observing a choice, we cannot tell, *based on this procedure and evidence alone*, whether he is indifferent between the two, or still choosing. If the recovery of preferences from observed choice is to be possible, a situation such as this cannot happen, i.e., we must have a procedure (which may be a conjunction of simpler procedures) from which it follows, in a finite amount of time, whether *or not* Edgar prefers A to B, and whether *or not* he is indifferent between A and B. As Ian Little (1949, p. 92) put it, "how long must a person dither before he is called indifferent?"

For Little, an important contributor to early revealed preference literature, undecidabilities are a good reason to want to cast preferences and indifferences aside.

This is an application of Pareto's methodological insight: if you cannot make sense of a notion using all the acceptable evidence and by appealing to reasonable procedures, reject the notion. Little's (1949, p. 92) argument emphasizes that it is market behavior that is the object of economics, not motives or even choice in general. Inconsistent preferences need not imply inconsistent market choices, meaning by consistency transitivity, non-satiation, and constancy (p. 91). Little argues that an agent may be irrational as per her answers to questions asked by an economist, or even perhaps as per her behavior in a controlled experimental, or other non-market situations, but still have a consistent observable choice behavior (purchases): "it must be said that there is something wrong with economics if it cannot explain perfectly consistent market behavior" (p. 93).

So a new theory of consumer behavior that does not depend on the dubious notion of preference, but relies instead on market choices and data, would seem desirable. Why bother with all the difficulties, and probable falsehood, of theories of choice in general when developing a theory of consumer behavior, as opposed to reducing the scope of the theory to actual choices in a market setup? This way we not only avoid having to struggle with the difficulties raised in the previous paragraphs, but we can also insulate at least this branch of economics from cognitive psychology and philosophy.

5.2.1.2. Samuelson's Weak Axiom of Revealed Preference

Pursuing a theory of consumer behavior 'freed from any vestigial traces of the utility concept' (p. 71), Samuelson (1938b) originally proposed three postulates:

The first postulate is that there is a demand *function*, i.e., a single quantity of each good *i* that the consumer purchases for each vector of prices, *p*, and income, *I*: $x(p, I)$ Further, the function is such that Walras' law holds for all $(p, I) : p' x(p, I) = I$ The second postulate is the homogeneity of degree zero of the demand function: $\forall t > 0, x(tp, tI) = x(p, I)$ for all (p, I), i.e., that only relative prices matter - if prices and income suffer the same positive, purely linear transformation, the quantity

purchased remains the same. The final, and crucial postulate, is what later became known as the weak axiom of revealed preference (WARP)[16]:

(WARP) $\sum_{1}^{n} p_i x_i(p', I') \leq \sum_{1}^{n} p_i x_i(p, I) \rightarrow \sum_{1}^{n} p'_i\, x_i(p, I) > \sum_{1}^{n} p'_i\, x_i(p', I')$ [17].

The proposed interpretation of this axiom scheme is the following: when a consumer purchases a bundle $x(p, I)$ of goods, facing prices *p*, and spending all her income *I*, when another bundle $x(p', I')$ is also affordable, then if the prices and income are such that the latter quantity is purchased, $x(p, I)$ cannot be affordable[18].

The appeal of the WARP is that it, supposedly, only uses respectable, easily observable notions. There seems to be nothing mysterious about prices: just jot down the number in the tag, incomes: just inspect the wallet, or goods purchased: just check the carrier bag[19].

[16] Actually, the first two postulates could be dropped because, if we accept to go along with Samuelson, they are implied by the WARP. See Samuelson ([1947] 1983, pp. 111-2) or Samuelson (1938c) for proof.
[17] A formal point should be noted. What is being called axioms are really open formulas, axiom schemes. They could be transformed into three axioms through universal quantification over x.
[18] "If this cost is less than or equal to the actual expenditure in the first period when the first batch of goods was actually bought, then it means that the individual *could have* purchased the second batch of goods with the price and income of the first situation, but did not *choose* to do so. That is, the first batch (x) was *selected over* (x')." (p. 65) Notice that here Samuelson is using a counterfactual (could have) without any qualms, which is interesting because, as has been mentioned in 4.6., the intensional logic of counterfactuals is far from transparent, and, certainly, the truth of a counterfactual cannot easily be determined with simple observations of choice behavior. Indeed, the individual could have purchased the second batch only if, among other conditions, he had been aware of that possibility. It is also noteworthy that Samuelson begins with 'choice' and ends with 'selected over'. Samuelson is worried with grounding consumption theory on "observable phenomena" but the translation of his axiom schemes into the metatheory smuggles many unobservables.
[19] "It will be noted that only prices and *quantities*, observable phenomena, appear in these expressions" Samuelson (1938a, p. 347 italics supplied). Samuelson's WARP, by implying Walras' law, does away with incomes. See Richter (1966, pp. 637-8) In the sense of just *looking*, it seems that prices, incomes and purchased goods are observable. We can see someone paying for a bag of goods, whose contents we may check. We can see how much this person earns per unit of time, or how much cash the person has available. But, of course, as argued in the previous chapter, just because we can observe these prices, these incomes, and these goods, it does not mean that we can observe the prices, incomes and goods that figure in the axioms. If we justify the axioms in the metatheory by showing that they are suitably related to intuitive notions of preference and belief, then the denotation we are giving to the terms in the language of the theory is not that of other homonymous terms 'prices', 'incomes', and 'purchased goods'.

5.2.1.3. Criticizing the Strong Approach

Does the WARP deliver the goods? First, there is the question of its status as a theory of consumer behavior. It does not take much thought to see that the WARP is implied by the classical, preference based theory of consumer behavior. It is the other direction of the implication that is interesting. Samuelson, in the original article, proved that the three postulates are equivalent to the compensated law of demand:

(eqs. 14 and 14.1 of Samuelson (1938b)) If $\sum_1^n p_i \Delta x_i = 0$ then $\sum_1^n \Delta p_i \Delta x_i \leq 0$

Using the language of differential calculus, this is the same as saying that the matrix of substitution effects (the Slutsky matrix) of $x(p, I)$ is negative-semidefinite. But Samuelson also showed that, when there are more than two goods, the postulates are insufficient to ensure that the Slutsky matrix is *symmetric*, as it necessarily is in the preference-based approach[20]. The WARP is thus weaker than the classical approach: there are demand functions satisfying the WARP that could not be the result of maximizing rational (transitive, complete, reflexive) preferences[21]. Since, according to Samuelson (1938b, p. 68);(1938a, p. 348);([1947] 1983, p. 116), the symmetry of the Slutsky matrix is an observable implication of the old approach, one might have expected him to try and strengthen his WARP. But that was not his take in Samuelson (1938b). For the Samuelson of this article, if the symmetry of the Slutsky matrix could

[20] For proof, see Mas-Colell, Whinston, & Green (1995, pp. 69-72), or Samuelson (1938a).

[21] It is transitivity that is sometimes lacking. Notice that with the WARP, if for some consumer unit we have a bundle A revealed preferred to a bundle B, and bundle B is revealed preferred to a bundle C, the WARP alone need not imply that A is revealed preferred to C. A concrete example, due to Hicks and taken from Mas-Colell et al. (1995, p. 35) is the following: take a consumer with 8 monetary units to spend on three goods. Assume that we observe $p = (2,2,1)$, $p' = (2,1,2)$ and the consumer buying $x(p,8) = (2,1,2)$, and $x(p',8) = (1,2,2)$. We see that $p^T x(p,8) = 8$, and $p^T x(p',8) = 8$, hence that the bundle $x(p',8)$ was affordable at the price-income combination for which the consumer chose $x(p,8)$. If the WARP holds, $x(p,8)$ could not have been affordable at the price-income combination for which the consumer chose $x(p',8)$. Indeed, $p'^T x(p,8) = 9 > 8$. $x(p,8)$ has thus been revealed-preferred to $x(p',8)$. Now take a third bundle: $x(p'',8) = (2,2,1)$, chosen when $p'' = (1,2,2)$. Simple calculations show that $x(p',8)$ is revealed preferred to this third bundle which, in turn, is revealed preferred to $x(p,8)$. So we obtained a cycle, which is impossible in the classical, preference based approach that assumes transitivity of preferences.

not be obtained from his WARP, so much the worse for that symmetry, and for integrability that necessarily goes by the board with it[22].

In fact, Samuelson (1938b, p. 68) writes that he had doubted integrability to begin with: "I should strongly deny, however, that for a rational and consistent individual integrability is implied, except possibly as a matter of circular definition." Moreover, he does not think that much observational content is being lost anyway, since symmetry could only be "subject to refutation under ideal observational conditions", and he has "little faith in any attempts to verify" it.

More interesting for us is whether the WARP successfully builds a theory of consumer behavior *independently of preferences*. When justifying his three postulates, Samuelson (1938b) originally appeals to "an idealized individual" (p. 63), and says that "the meaning of this [the WARP] is perfectly clear and will probably gain ready acquiescence[23]" (p. 65) He is, therefore, referring to an implicit standard shared with his readers. The reason why the WARP gains ready acquiescence is that it is implied by the original preference based approach which, in turn, is based on the simple, intuitive maxim that people tend to choose what they prefer. In other words, preferences end up making an appearance, and the plausibility of the new approach comes from the same intuitions that underlay the older approach[24].

We could, of course, think of other justifications. For instance, that the WARP offers a good fit with the data, is computationally economical, and predictively accurate. As far as I know, this sort of justification, which of course would take the WARP as just another candidate axiom scheme, is never seriously entertained. But even if these claims were true, such a justification had been offered and thus objections along the lines of 4.7. avoided, there would still be the matter as to *why* the WARP works. And any

[22] The problem of integrability is that of getting the utility function that underlies a given demand function. The symmetry of the Slutsky matrix is a necessary condition.
[23] I suppose that Samuelson meant that the WARP would gain ready acquiescence, not its meaning.
[24] This point was first made by Sen (1973, p. 57 italics supplied) who took it as showing the failure of the strong approach. "Faith in the axioms of revealed preference arises, therefore, not from empirical *verification*, but from the intuitive reasonableness of these axioms interpreted precisely in terms of preference."

explanation of this would involve the *truth* of instances of the WARP, suitably translated into some metatheory where preferences will make an appearance.

I thus conclude that the WARP has not been justified without an appeal to the same intuitions that underlie the preference based approach. All it does is brush preferences under the metatheoretical carpet. In this sense, it does get the *theory* rid of preferences, but not economics. Without the preexisting intuitions surrounding preferences, we would have no reason to develop these axioms to begin with.

5.2.2. The Weak Project

We have seen that we need the same intuitions underlying the preference based approach to justify the WARP, and that the latter is (empirically) weaker than the former: it does not imply all the "meaningful statements" of the former. The goal of the weak version of early revealed preference theory was to strengthen the WARP to do just this. It no longer tries to get rid of the illegitimate notions of utility and preference, but merely to derive sentences, in the theory, that capture *all* the observable *market* implications of the classical approach, without using preferences or utilities[25].

This was achieved with Hendrik Houthakker's (1950) strong axiom of revealed preference:

(SARP) If for every t and T such that $t \leq T \in \aleph$, $p^{t-1}x(p^t, I^t) \leq p^{t-1}x(p^{t-1}, I^{t-1})$, and for some m, n such that $0 \leq m < n \leq T$, $x(p^m, I^m) \neq x(p^n, I^n)$, then $p^T x(p^T, I^T) < p^T x(p^0, I^0)$

In other words, if we have a finite list of bundles such that the first is revealed preferred by some consumer to the second, the second to the third, etc., then the last one is *not* revealed preferred to the first. The SARP is also clearly implied by any theory of consumer behavior that is based on transitive and complete preferences. But, more

[25] See Houthakker (1950, p. 161): "The main object of our investigation is to find a proposition which, apart from continuity assumptions, summarises the entire theory of the standard case of consumer's behavior. (…) Such a proposition should imply and be derivable from utility analysis."

importantly, Houthakker (1950) proves that the SARP is, unlike the WARP, sufficient to ensure that there is some rational preference relation rationalizing whatever demand function observes it. In other words, the SARP is equivalent to a theory of consumer behavior based on rational preferences, but its only variables are incomes, prices, and quantities purchased. So each SARP can really be thought of as a protocol for classical consumer theory using only the evidential basis that the revealed preference theorists condone.

But there is an interesting snag at this juncture: is the SARP really *necessary* for the weak project? The reason for this question is the following: I have shown that the only plausibility the WARP or the SARP enjoy comes from a metatheoretical justification that appeals to our preexisting intuitions about preference. But then, since we are committed to this notion anyway, we are as much justified in assuming that the WARP holds in a market setting, as we are in assuming that it holds, suitably rewritten, for *all* choices, not just market ones. If we do assume this, then Sen (1973, pp. 58-9) tells us that the SARP is implied by the WARP and therefore that the WARP and the SARP are equivalent.

Hence, if the SARP gives all the empirical implications of the old consumer theory, then so does the WARP. The SARP is, after all, useless! As Sen (1973, p. 59) puts it: "[insulation of market choices from choice in general] is how in the theory of consumer's behavior, the man can get away with satisfying the Weak Axiom over all the cases in which his behavior can be observed in the market and nevertheless harbor an intransitive preference relation."

I believe this criticism misses the most important contribution of what I am calling the weak project[26]. The weak project derives protocol sentences for the theory of consumer behavior using a specific evidential basis. The WARP exhausts the observational implications relative to this basis (market prices, incomes and quantities purchased) if you assume that it holds for all other choices, too. But to *express* this assumption, you need further notions, viz. those necessary to express non-market choice. If you do not want to introduce these notions into your *theory*, then you need the SARP to get all the

[26] Even disregarding important difficulties, to be discussed in the next section, with assuming that the WARP holds for all choices.

observational implications, even if you use these notions in your *metatheory* in order to justify the interest in the SARP[27]. The raison-d'être of the revealed preference approach is in its limitation to a strict evidential basis, including only market-based observations, and disregarding verbal protocols and other non-market evidence.[28]

It is better to take early revealed preference theory as developing consumer theory in an empirically conscious language, amenable to the creation of observation sentences for a limited evidential basis. The fact that the theorists were working with an admittedly small evidential basis could but need not be construed as undue *prejudice* on their part, but as a way of sidestepping serious difficulties that more inclusive bases entail.

5.2.3. Summary

Our criticism of the early revealed preference theories should not blind us either to their contribution or to the lessons that their failure supplies. We have seen the danger of trying to simply get over our intuitions. They may end up, as in Samuelson, simply going undercover, and we may thus lose control over their operation[29]; or else we may so change our games that we fall prey to the sort of criticism discussed in 4.7.: of saying

[27] Cf. Mas-Colell (1982, p. 75): "The essence of revealed preference theory is the realization that the observable choice data will be far from inclusive of all conceivable choice experiments."
[28] Sen (1973, pp. 59-60) seems to grant the point, but he believes that the revealed preference position is, in general, misguided, even from its own standpoint: the axioms are not 'verifiable' (Sen's word, p. 59-60), as Sen thinks revealed preference theorists mean them to be. They are not verifiable because "to check whether the Weak Axiom holds for the entire field of market choices, we have to observe the person's choices under infinitely many price-income configurations" p. 56. As far as I am aware, revealed preference theorists are never very articulate in their methodological reflections. They usually make passing remarks on "observable implications" but neither subscribe to any particular school, barring Samuelson's occasional references to operationalism, nor defend their views by an appeal to explicit doctrines like verificationism, much less to the strong verificationism that Sen refers. Their overall position, ambiguities and all, is well summarized in the following passage, found in Samuelson (1938a, p. 344): "A meaningless theory (…) is one which has no empirical implications by which it could conceivably be refuted under ideal empirical conditions." This is as vague as it could get, of course: 'could', 'conceivably', 'refuted', 'ideal conditions' all need spelling out.
[29] It is almost ironic that Samuelson ([1947] 1983, p. xvi) should write that "working scientists, to tell the simple truth, have neither the time nor the patience to bother with the history of their subject: they want to get on with making that history. Philosophers of science, historians of science, sociologists of science, may not be without honor in their own houses; but the customers who take in their washings, and swap garments with them, are unlikely to be working scientists still in the prime of life." The methodologist and the historian of thought are always the ones who have the last laugh.

a lot about nothing that interests us. I have also tried to leave clear the importance of the distinction between theory and metatheory.

Moreover, in spite of my misgivings as to the appropriateness of Sen's criticism of the weak project, it nevertheless makes clear the potential difficulties of artificially insulating theories of choice to very particular kinds of choice. It thus hints at a way forward. If the weak axiom is assumed for *every* choice, it implies the strong axiom which, we are told, is equivalent to the preference based approach. An expectation is thus induced that if we enlarged our evidential basis to general choice behavior, we could "behaviorally characterize" preferences. Fortunately, much subsequent work has tried to generalize revealed preference theory to choice in general. The next section will try to make progress by looking at this literature.

5.3. Revealed Preference Theory

5.3.1. Marcel Richter's Framework

It was Marcel Richter (1966), (1971) who first presented Revealed Preference Theory in its most general formulation. As Mas-Colell (1982, p. 76) writes, "it was M. Richter who recognized the revealed preference theorem for what it is – a theorem in set theory."

The framework Richter proposes is simple, deceptively so. First, there is a universe of alternatives of action *X*, for instance, $X_{example} = \{go\ swimming, read\ a\ book, go\ to\ a\ spa\}$. The nonempty subsets of *X* are translated as the budgets. With our $X_{example}$ there are at most seven budgets[30], an example of which is $B_1 = \{go\ swimming, read\ a\ book\}$. Richter defines choice as a set *function*, $h(B)$, with domain a selection of budgets β, mapping each budget in β to one of the budget's subs*ets*, which may be the empty set. In other

[30] 2^3 minus one for the empty set.

words, it is a function selecting from each budget one or more alternatives[31]. The formula $x \in h(B)$ is translated by Richter as meaning that 'x' is chosen, and the formula $y \in B$ as meaning that 'y' *could have been* chosen. Pursuing our example, if $h(B_1) = \{go\ swimming\}$ then 'go swimming' is chosen, and 'read a book' could have been chosen.

Preferences, as usual, are defined as a binary relation over $X : x \succcurlyeq y$ means that alternative 'x' is liked at least as much as alternative 'y'. Richter distinguishes preferences *simpliciter*, i.e. any two-place relation over X, from what he (1971) calls *regular* preferences, which are complete, reflexive and transitive two-place relations over X [32]. It is noteworthy that he defines a choice as rational if *some* preference rationalizes it, in other words, if the elements chosen for each budget are the most preferred elements in that budget according to some preference, even if the preference is not regular. This definition is interesting because it implies that a choice is rational even if it is only rationalizable by an intransitive preference relation[33].

Richter's contribution is, however, in his attempt at offering us necessary and sufficient *behavioral* conditions for choice to be rationalizable *simpliciter*, and to be rationalizable by regular preferences.

5.3.1.1. Richter's First Axiom

The first condition is given by the following axiom scheme, which I will call *Richter's First Axiom*:

RA1: $\forall_{B \in \beta} B\ h(B) = \left\{x : x \in B \,\&\, \forall_{y \in B} y\ \exists_{M \in \beta} M\,(x \in h(M) \,\&\, y \in M)\right\}$

[31] Richter allows there to be more than one chosen alternative from a given budget: $h(B)$ need not be a singleton. This tries to preserve relations of indifference but Richter is silent about the problem I discussed in 5.2.1.1. on behaviorally deciding indifference.
[32] In other words, those to which economists, Richter (1966) included, ordinarily vouchsafe the encomium "rational".
[33] The question then arises whether *any* choice is irrational. Richter (1971 Theorem 1) gives an example of an irrational choice.

RA1 is really a generalized WARP. It is satisfied by a choice function that chooses an alternative for some budget if and only if that alternative is chosen over each of the other alternatives in the budget for some budget in the set of permissible budgets (β). Richter shows that if this condition is observed by a choice function, then it can be rationalized by some preference relation $\succcurlyeq$. Importantly for us, Richter (1971, p. 33) teaches us how to construct such a relation: $x \succcurlyeq y \leftrightarrow \exists_{B\in\beta} B(x \in h(B) \,\&\, y \in B)$.

Translated into the metalanguage: an alternative is liked at least as much as another if and only if it is chosen over that alternative for some permissible budget.

I reiterate that not every conceivable budget need be permissible, i.e., in β . The budgets were defined as nonempty subsets of the set of possible alternatives, but Richter leaves open whether choice is over *every* budget or merely from a selection of these[34]. This is crucial for the behavioral credentials of the revealed preference project. We seldom observe people choosing from every possible combination of alternatives. Instead, we observe their behavior in a few, concrete, dated choice situations. What Richter's First Axiom tells us is: given the observations we have of someone's behavior (choices) at several circumstances (budgets), could there be *some* preference relation that is such that every choice is the best alternative in the budget? RA1 gives us the necessary and sufficient conditions for that to be so.

5.3.1.2. Richter's Second Axiom

Our discussion of Sen's criticism of the SARP in 4.2.2. should, however, make us expect that this generality over permissible budgets comes at a cost: RA1 does *not* give us sufficient conditions for a choice to be rationalized by a transitive preference. For that, we need a stronger axiom scheme, generalizing the SARP, which we name *Richter's Second Axiom*:

[34] He also leaves open, unless technically necessary, whether choice is always *decisive:* whether at least one alternative is chosen from every permissible budget. Actually, RA1 is also necessary and sufficient, if choice is decisive, for choice to be rationalized by *reflexive* preferences. Cf. Richter (1971, p. 34) Richter does not give a behavioural criterion of decidability. All he says is that, if only one chosen alternative is allowed, then "an element is selected [from *h(B)*] by some random device." (p. 31)

RA2: $\forall_{B\in\beta} B\ h(B) = \{x : x \in B \,\&\, \forall_{y\in B} y\ \ xWy\}$

xWy stands for a simple yet tediously long formula requiring the existence of a sequence, of whatever finite length, of alternatives, starting with x and ending with y, such that each element in the sequence is chosen, for some budget, over the element immediately following it in the sequence.

If RA2 is satisfied by a choice function, then choice can be rationalized by a transitive, *reflexive* preference relation: $x \succcurlyeq y \leftrightarrow xWy \vee x = y$ (Richter (1971, p. 35)). Not just this, if we are willing to assume that the choice is decisive, i.e., that only one alternative is chosen for each budget, then if it observes RA2, it can be rationalized by transitive, reflexive, and *complete* preferences[35]. Hence, Richter (1966, p. 639) writes that it "emerges as the behavioral, revealed preference condition which is equivalent to rationality."[36] Also, Richter (1971, p. 30) believed that this was a formalization of the "intuitive notion of rational behavior."

Clearly, the preference *relation,* regular or not, obtained from Richter's procedures, need not be the only preference relation that the evidence justifies, i.e., that would rationalize observed choice behavior.

This is easy to show with a simple example. Suppose we have observed Jane buying a dress, a pair of jeans, and a blouse. Suppose also that the dress was bought at a store that only sold *n* different dresses, the jeans at a store that only sold *m* different jeans, and the blouse at a store that only sold *p* different blouses. We thus have three budgets, and have observed three choices: $B_1 = \{dress_1, dress_2, ..., dress_n\}$, $B_2 = \{jeans_1, jeans_2, ..., jeans_m\}$, $B_3 = \{blouse_1, blouse_2, ..., blouse_p\}$, $h(B_1) = \{dress_a\}$, $h(B_2) = \{jeans_b\}$, and $h(B_3) = \{blouse_c\}$.

[35] That there is a regular preference relation extending an incomplete, transitive one can be shown as a matter of course from the axiom of choice and known equivalences.
[36] Actually, he writes this of another axiom, equivalent to RA2 for decisive choices.

Clearly, this choice function satisfies RA2. If we construe a preference relation using Richter's procedure, we obtain the following incomplete, transitive and reflexive preference relation:

a) each chosen garment is indifferent to itself and preferred to each one of the other garments of the same kind, but is not comparable with any garment of a different kind.
b) garments that are not chosen are indifferent to themselves and less preferred than the chosen garment of the same kind. They are not comparable with any other alternative.

It is easy to see that there are many ways to fill the gaps and extend this preference relation to a regular preference that rationalizes this same choice. What is crucial to notice is that these extensions may contradict each other: e.g. one may put $jeans_b \succ blouse_c$, whilst another extension puts $jeans_b \preccurlyeq blouse_c$. In other words, it is possible that contradictory preference relations are compatible with all the behavioral evidence we *do* have. Naturally, *if* we could put the acting subject to the test, in our example by forcing a choice over the budget $B_4 = \{jeans_b, blouse_c\}$, for instance, *then* one of the preference relations would have to be discarded[37].

Notice, however, that it is conceivable that the budgets needed to increase determinacy may not even exist in the actual world, past, present or future, let alone be observable. If the required observations over further budgets necessary to eliminate all the contradictory regular preference relations are unavailable, either practically or as a matter of principle, is there any sense to saying that one of two contradictory preference relations that nevertheless rationalize all the present and future available evidence equally well must be the right one? I will deal with this question, in its full generality, in 5.4..

[37] Or both, if the observed extended choice function turned out *not* to be rationalizable by regular preferences.

5.3.2. Assessing Revealed Preference Theory

When I discussed Pareto's work in 5.1., I showed that many utility functions can equally reproduce the behaviorally significant properties of preferences. In particular, Pareto does not believe that any evidential basis could ever determine preferences to an extent that would justify a utility more definite than some positive monotonic transformation of its differences. Unfortunately, he simply assumes that preferences could be obtained from people's behavior, without showing how.

Richter, in his turn, teaches us how to construct preference relations from our observations of an agent's choices under several circumstances, observations which are summarized in a choice function over budgets. If Richter's proposals are sound, it seems that we have made progress: if we can observe an agent choosing, we can recover her preferences, if any there are to be recovered. Since I argued in 4.3. that action is caused by coming to form an appropriate reason, i.e., desires and beliefs, it seems that half of our problem is solved. Unfortunately, revealed preference theory suffers from severe difficulties if it is to serve for these purposes. I discuss them in the next two subsections.

5.3.2.1. Problems with counterfactuals

One important problem with the framework is with its underlying logic: it illustrates the difficulties discussed in 4.6..

Richter's apparatus is only empirically interesting if the choice function ranges over more than one budget: if it ranged over none or over just one budget, then the axioms would be trivially satisfied. Succinctly, any one isolated choice can be rationalized by complete, transitive and reflexive preferences.

Notice, however, that the axiom schemes range over the choice *function*. But it is obvious that we can only observe one choice at a time. We could try to translate the choice function as a conjunction of material conditionals, written in the present tense: *if*

the agent faces budget B_1, *then* it chooses $h(B_1)$, *and if* it faces budget B_2, *then* it chooses $h(B_2)$, *and* …, *and if* it faces budget B_n, *then* it chooses $h(B_n)$. But this does not seem right. At a moment in time, the agent can only face, and be observed to face, one budget. Let us say that it faces budget B_m, and that it chooses $h(B_m)$. Then, at this moment, the conjunction of conditionals is *true*[38], but so is the conjunction obtained by keeping the antecedents the same and denying any of the consequents, except the m^{th}. In other words, thus translated, the axioms are as trivially satisfied as if the choice function ranged over just one budget set.

In order to make the theory interesting, we could try to translate it along the same lines that we have seen Richter translate each *h(B):* if the agent *were to face* budget B_1, then it *would choose* $h(B_1)$, and if it *were to face* budget B_2, then it *would choose* $h(B_2)$, and …, and if it *were to face* budget B_n, then it *would choose* $h(B_n)$. Thus translated, we cannot conclude for the truth of the conditionals simply because the antecedent is false, i.e. because the agent does not face budget B_i. Whereas a material conditional, a sentence of the form $\phi \rightarrow \alpha$, is true if and only if either the consequent (α) is true or the antecedent (ϕ) is false, the truth of a counterfactual conditional involve a reference to the possible world most similar to the actual world for which the antecedent is true (see Lewis ([1973] 2001)). Thus understood, the choice function does not involve just what the agent *actually* does, but what the agent *would do* in specific possible worlds where it faced budget B_i. The problem, however, is on determining these dispositions with the behavioral evidence allowed. Also, the first-order logic that Richter is using in his theory does not reproduce the inferential relations that are smuggled by this translation into the metatheory.

Finally, we might try to avoid counterfactuals, and interpret choice functions as referring to the past: the *if the agent were to face* becomes *when the agent faced* and the *would choose* becomes *chose.* We could, then, follow Richter's procedure to recover a

[38] The reason for this is that if the sentence 'ϕ' is false, '$\phi \rightarrow \alpha$' is true, whatever the truth value of sentence 'α' and every sentence 'it faces budget B_n" for $n \neq m$ is false.

preference relation rationalizing the choices *actually* observed. But here we face another problem.

The problem is with what the preferences thus recovered are supposed to be good for. Thus translated, the preferences we recover apply to observed choices made in the past. What justification do we have to project them onto the future? We want to use whatever preferences we recover to explain other, thus far unobserved, choice behavior. But, as Joan Robinson (1962, p. 49) notices, observations of behavior are in real time, and so any evidence against the axioms could be rationalized as evidence for a change in tastes[39]. In other words, preferences either have to be *assumed* constant, or else we need to offer behavioral criteria of their change.

It is noteworthy, however, that we do not usually have any particular trouble deciding whether other people changed their minds about something, *using nothing more than behavioral evidence*. Yet this is a source of trouble for the revealed preference theorist because of her disregard for verbal behavior.

5.3.2.2. Problems with the determination of propositional content

Richter's framework has two primitives. The choice function, and the set *X*. It is assumed that we can observe *both.* But is this assumption tenable?

I do think it is plausible to assume that we can observe *that* an agent chooses: I think we can usually tell if people are *doing something*, as opposed to merely moving[40]. But can we observe *what* a subject chose? Again, in a sense, we can: when we observed Jane acting, we saw her go to the table, put a slice of cake on a plate, and proceed to ingest it. We can look at the table and jot down the available options. We thus seem to have

[39] This objection, as Robinson admits, is old, and the people in the revealed preference tradition were certainly aware of it. See, e.g. Samuelson (1938a, p. 346) or Little (1949, p. 96).
[40] Naturally, this social skill, unless it is *also* behaviourally characterized, is a death knell to the claim that revealed preference "behaviourally characterizes" rational behavior, in any strict sense of 'characterizes'. Still behavioural reduction is *not* my goal. In fact it could not be since it is impossible.

determined one budget: the set of foods on the table, composed of a carrot cake and a chocolate cake; and one selection over the budget: a slice of chocolate cake[41].

But has she chosen the chocolate over the carrot cake? Might she not have chosen the brown over the orange cake? Or the only gluten-free cake on the table? Or the homemade cake that her best-friend brought to the party?

There are many true descriptions of what Jane did[42], not all relevant to explain or understand her behavior. If Jane chose to get a slice of chocolate cake because she knew her best-friend had made it, and she wanted to be nice to her best-friend, it is plausible that if her best-friend had brought the carrot cake, she would have chosen the carrot instead of the chocolate cake. If we define X as a set (cf. 4.6), and the budget as $\{chocolate\ cake, carrot\ cake\}$, it is obvious that the axioms would tell us that Jane's behavior is not rational.

We could try to come up with solutions to this problem without resorting to verbal reports. We might, for instance, try to increase the number of observations, judiciously setting up the circumstances under which Jane chooses in order to determine her space of choice. For instance, we could try to have her choose a homemade, gluten-free chocolate cake against a homemade, gluten-free orange cake to see whether it was a matter of chocolate versus orange. But this suggestion is problematic.

Indeed, for this suggestion to work, we would require a list of characteristics that might be relevant *to Jane*. Where does this list come from? We could try to list characteristics that *we* believe to be relevant. But where does our judgment of relevance come from? What is salient to Jane? What inductive paths does she follow? If we have to assume all this, it is dubious what sort of insight the framework brings. Moreover, this suggestion is suspect in that ordinary people do not follow it at all. When they are in doubt, they

[41] Which, of course, need not be unique because the set of alternatives chosen from a budget need not be a singleton. I will ignore these issues with decidability which are, as seen in 5.2.1.1., the result of ignoring verbal reports.

[42] Hopefully, the reader will agree that the descriptions we chose are not merely possible, but plausible: people *do* choose cakes for their main ingredient, for their colour, to suit food-allergies, and because a dear one has made or recommended it. These are not highly idiosyncratic descriptions that a revealed preference theorist might brush aside as philosophical quibbling. We *can* have a legitimate doubt whether the reason why someone chose a particular cake is because it is a chocolate cake, or a gluten-free cake, or the most luscious-looking cake, etc.

ask each other questions. In our case, we would do well to ask Jane what exactly she was doing and why. The cost of ignoring verbal reports keeps rising, although it should by no means be thought that the need for this sort of tentative speculation would be eliminated if they were incorporated.

It is also noteworthy that this framework, as it stands, does not reproduce the calculus we usually employ when determining other people's tastes. It is likely that we would infer from our knowledge that Jane loves, say, chocolate, that she also loves chocolate *cake*. We may never have observed her having cake, but we may have the prior belief that people who love chocolate *tend* to love chocolate cake. Indeed, we not only observe choice over a small number of budgets and extend the preferences thus recovered, but we also infer about alternatives we have never seen chosen or chosen over. Moreover, we usually pay attention to time frames. At 8 p.m. we may much prefer a hearty meal to a nap. But at 9 p.m., after gorging at dinner, what we really want is a nap. We might explain this by saying that we are in the presence of two different goods: a hearty meal after another hearty meal is obviously a different thing than a hearty meal when hungry. It depends on the modulus: on how we discriminate alternatives, over time, place or whatever dimension theory tells us is relevant. But how is this determined using the evidence allowed?

The criticism in this subsection points to the fact that *X* is as much in need of behavioral characterization as preferences. There is thus no presumption that it will be in any way determinate, as preferences themselves were shown not to be[43]. It is also clear that it should not be seen as a set in the mathematician's sense (cf. 4.7.). The closed form of revealed preference theory leaves no room for the speculations, analytical hypotheses, and tentative projections that we need to make to determine the propositional content of agents' attitudes.

[43] This is of a piece with Quine's thesis of the inscrutability of reference. See Quine (1968).

5.3.3. Summary and Conclusions

Based on the choice behavior we observe, and the assumed observable circumstances in which it takes place, can we *construct* an agents' preferences? If yes, what preference or preferences are those? Given preferences with certain properties, what kind of constraints on behavior do they pose? This is what Richter's axioms try to give us[44].

However, what an agent prefers only gives us half the reason for her actions, the half connected with desires. As I argued, reasons also involve beliefs. Indeed, I may much prefer to go to the cinema than staying home. Yet, if I believe that the cinema is closed, I will stay home. It would be wrong to say that I reveal a preference for staying home over going to the cinema. Decision theorists have therefore paid much attention to the problem of recovering both beliefs and desires from behavioural evidence. It was a procedure developed by Frank P. Ramsey (1926) that first taught us how to do this. To control for beliefs and to get preferences, all we need is to find a proposition an agent finds as likely to be true as false, what Ramsey calls an "ethically neutral proposition" (p. 73)). Asking the agent to choose between gambles whose outcome depends on the truth of such a proposition allows us to recover the agent's preferences over outcomes (see p. 75 for details). Once we have determined his preferences, we can control for these, and recover degrees of belief (see pp. 75ff). We represent degree of belief with probabilities, the same way we used utilities to represent preferences. In economics, Leonard Savage's (1972) axiomatization of subjective expected utility theory is the *locus classicus* of this sort of approach.

It is noteworthy that many of the contributors to decision theory defend their work with an appeal to intuition and common-sense. Savage (1972), for instance, writes that he "hope[s] that the notion of probability defined here is consistent with ordinary usage" (p. 27). Maurice Allais (1953), when he first presented his eponymous paradox, had as a goal to show that Savage's axiomatization did not explicate our common-sense notion of rationality. He writes that there are people who anyone would count as rational but who violate Savage's independence axiom (cf. p. 525 and p. 530). John von Neumann

[44] Or so I am construing them. Implementation is never discussed by Richter in the articles quoted.

and Oskar Morgenstern ([1944] 2007, p. 361) follow suit: "if the mathematical proof fails to establish the common sense result, then there is a strong case for rejecting the theory altogether."

The sort of problems that decision theory suffers from, if it is to serve my purposes, is similar to those diagnosed above for revealed preference theory. The inferences allowed are, sometimes, too strict, other times, too loose. In some instances we can reasonably conclude more than these theories allow: for instance that Jane likes chocolate when we see her eating a chocolate cake[45]. In other instances we are not justified in concluding as much as these theories seem to imply: if yesterday Jane chose the chocolate cake, it is a stretch to conclude that she either is irrational or will *always* choose the same dessert when confronted with the same menu, even if her beliefs do not change.

I should, however, add that decision-theorists have been less reluctant to allow verbal reports into their evidential basis[46]. Still, as argued in 5.2.1.1., the introduction of verbal reports requires a theory. Indeed, what people mean by what they say is as much and as little transparent as what they like or believe. We determine everything together as we interact and make sense of them: we are always radically interpreting, even if more often than not homonymous translations of their sentences into ours works well[47]. To close the decision theoretical circle would require, as Davidson (1980, p. 154) urges, that we recover beliefs, desires *and meanings*, all three together, from an acceptable evidential basis, using sound procedures. Quine's (1964) work on radical interpretation is a step towards a theory of meaning. Davidson (1965);(1967b) puts Quine's work in a decision-theoretic garb by developing a theory of meaning by way of Alfred Tarski's (1936) work on truth. Davidson (1980) then develops what he calls a unified theory of

[45] For an overview of nonmonotonic logic, see Makinson (2005).
[46] Sometimes with difficulty. In Ramsey's case, for instance, for the procedure to work the agent must believe that the wager being offered is real. Ramsey (1926, p. 72) nonfacetiously writes that "if then we had the power of the Almighty, and could persuade our subject of our power, we could, by offering him options, discover how he placed in order of merit all possible courses of the world." For an alternative theory not assuming this, see Jeffrey (1983, pp. 157-8).
[47] Cf. Davidson (1973b, p. 125): "all understanding of the speech of another involves radical interpretation." Cf. Quine (1968, p. 46): "we will construe a neighbor's word heterophonically now and again if thereby we see our way to making his message less absurd." See also Quine (1964, pp. 59, 78 and 203).

thought, meaning and action by incorporating his theory of meaning into Richard Jeffrey's (1983) logic of decision[48].

In this literature, Hayek's intuition that we need to assume that we share a lot with other people is vindicated. Whenever any author tries to spell out the sort of tentative projections, hypothesizing and speculation that is necessary to produce an alternative theory of another agent, they always conclude for the necessity of assuming that we are all birds of a feather, and that other agents see the world the same way we do. As Davidson (1997, p. 140) writes, "before there can be learning there must be unlearned modes of generalization. Before there can be language there must be shared modes of generalization."

In the next section, I fulfil two promises. First, that of further discussing to what extent of certain properties like transitivity are constitutive of notions such as preferences, and second that of studying the extent to which our attributions of preference and belief are indeterminate, and the implications of such indeterminacy.

5.4. Independence and Indeterminacy

As noticed in 4.6., there are properties and interrelations that are often said to be constitutive of certain notions, for instance that preferences are transitive, or that resources are scarce. Another example are the axioms of independence, typical of decision theory in its expected utility variety. These axioms tell you that if you mix lotteries in a certain way with *another* lottery, the lotteries thus resulting should be ranked the same way as the original, before the mixing, were. Whereas with two goods we may often have *both*, and thus relations of complementarity and substitutivity make sense, with mutually exclusive outcomes either one or the other comes to pass: there is no justification for contaminating your preferences when one results with what your preferences would be if another resulted. So if you prefer to go for a walk than staying

[48] For a discussion of Jeffrey's work, see Broome (1990).

home if it does not rain, and are indifferent between the two if it *does* rain, then you should go for a walk!

Yet, it is said that agents frequently violate these axioms (cf. Tversky & Kahneman (1992, p. 297) or Savage (1972, pp. 100-101)). Allais (1953) was one of the first to notice this. He concludes, as seen above, not that people are irrational but that an axiomatization implying independence does not explicate rationality. Still, it is more usual to say that the axiom *is* implied by our intuitive notions of rationality, but that agents do not conform to the ideal (See, e.g. Tversky & Kahneman (1986)).

What is interesting about violations of these axioms is that we are often able to rationalize them away. As Savage (1972, p. 101) notices, "the reinterpretation needed to reconcile all sorts of behaviour with the utility theory is sometimes quite acceptable." Maybe the agent has a preference for the gamble itself, maybe we should refine the underlying space of prizes, as opposed to rejecting the truth of the axiom. Samuelson (1952, p. 677) writes that we are "caught between the Scylla and Charybdis of theoretical formulation and operational empirical hypothesis formation." For him, the move of changing the underlying space to suit the axioms must have a limit. "If every time you find my axiom falsified, I tell you to go to a space of still higher dimensions, you can legitimately regard my theories as irrefutable and meaningless" (p. 677).

But I believe that there is a more important lesson to be learned. The important lesson is not that we can always redefine our terms so that the falsest of theories comes out true. The interesting thing is that when we redefine the underlying set of options to deal with apparent violations of the axioms, the redefinitions are often plausible. What Samuelson's dilemma shows is that what is evidence of the violation of the axioms is as much evidence that we do not understand agent's beliefs[49].

It seems to me undeniable that everybody *sometimes* violates the independence axiom and every other decision-theoretical axiom. We all make mistakes. Sometimes we are tired or are not paying attention, other times we have no excuse and are just being foolish. What is important is that we be able to determine when we and other people

[49] Interestingly, Savage (1972, p. 25) approvingly refers to Quine (1951), an article where the latter questions any firm distinction between the analytic and the synthetic. See 4.4.2.

make mistakes, and to account for them. In some instances, it may be more plausible to say that a person's behaviour violates the axioms, and in others that the baseline set of alternatives needs to be altered. But where we draw the line will depend on a judicious evaluation of the concrete circumstances. Only if we want a closed theory do we face a dilemma.

In order to try to better describe behaviour, and capture the pattern of the departures from the baseline expected utility case, economists have been adopting ever more complex axiomatizations. Kahneman & Tversky's (1979) Prospect Theory is a case in point[50]. Authors in this vein try to have their cake and eat it too: the new axioms are not intuitive, yet they are a departure *from* clear intuitions, ever present. But these theories face a significant problem from the perspective I have been adopting: they seem to be far from reproducing our everyday ways of understanding other people. Naturally, this does not impugn their potential interest for *other* projects.

I believe that work in the field of pragmatics indicates an alternative to complex axiomatizations. Suppose someone tells you that Jane, whom you both know to have taken a long holiday in the Caribbean, is as red as a lobster. Would you think that your interlocutor was trying to say that Jane was *literally* as red as a lobster? It is highly unlikely: it is *obvious* that Jane's skin is not as red as that of a lobster. In fact, it is obvious to you that it is obvious to your interlocutor that it is obvious to you etc. that Jane's skin is *not* as red as a lobster. But, if you have no reason to believe that your interlocutor is being uncooperative, you search your common-ground for hints as to what she might mean. Jane's holiday is probably salient to both of you, and everybody knows that lobsters are red, so it is easy for you to *get* what your interlocutor was conveying: that Jane is very sunburnt. Notice, however, that the sentence 'Jane is as red as a lobster' does not imply, or is implied by, the sentence 'Jane is very sunburnt'[51].

[50] See also Tversky & Kahneman (1986, (1992), and Wakker (2010).

[51] The reasoning I am following can be turned into an explicit argument by appealing to Paul Grice's (1989, Part 1) Cooperative Principle. The interlocutor is flouting Grice's maxim of quality, which entreats you to "try to make your contribution one that is true" (Grice (1975, p. 27)). But the disregard of your interlocutor for the maxim is so obvious that, since you have no reason to suspect lack of cooperation, you search for what she might be trying to convey *by flouting it*.

An even better example of the difference between sentence meaning and utterer's meaning is shown by studying the meaning of the logical connectives. Take the English construction 'if … then'. Munkres (2000, p. 7) writes that "in everyday English the meaning of the 'if … then' construction is ambiguous. It always means that if *P* is true, then *Q* is true also. Sometimes that is all it means; other times it means something more: that if *P* is false, *Q* must be false. (…) Mathematics cannot tolerate ambiguity, so a choice of meaning must be made." The first meaning is, of course, the same as that of the conditional in standard mathematics. Munkres (p. 7) gives the following example of an English sentence where the construction takes the second, stronger, meaning:

a) "Mr. Jones, if you get a grade below 70 on the final, you are going to flunk this course."

According to Munkres, "Mr. Jones knows that if he gets a grade below 70, he will flunk the course, but if he gets a grade of at least 70, he will pass" (p. 7). I suppose that Jones learnt this by hearing the professor utter sentence a).

But there is an alternative way of looking at a). Suppose Jones gets a grade of 75 and still flunks. When he confronts the professor, the latter tells Jones that he had to get a grade of at least 80 in order to pass. Now, the professor did not utter a falsehood: a) *is* true. We should accuse the professor of misleading his student, not of misusing language or even of *saying* something false[52]. It is obviously relevant to Jones whether he needed to get at least 70 or at least 80. In saying the former the professor was not being as informative as the case required. In a nutshell, either the professor meant by a) what Jones interpreted him as meaning, or the professor was not being cooperative when Jones thought he was.[53]

We thus have two possible approaches. One is that followed by Munkres, of enriching the semantics of the language, and enumerating all the possible meanings an expression

[52] We should also accuse the professor of lying. To lie is not so much to say a falsehood, but to represent yourself as believing something you do not.

[53] In this case, it is Grice's (1975, p. 26) maxims of Quantity that are being violated. These maxims entreat us to "1. Make [our] contributions as informative as is required (for the current purposes of the exchange). 2. [to] not make [our] contribution more informative than is required." Notice, however, that in the example of Jane's sunburn our interlocutor *was* being cooperative, and hence *flouting* the maxim of quality. In this example the professor was *not* being cooperative, and he *violated* the maxim of quality.

or a sentence might mean. The other is to assume one baseline literal meaning, and look to pragmatics, to the context of language use, and broad principles, to discover what the speaker was trying to convey. (Cf. Stalnaker (1987, p. 126))

Similarly, people will behave reasonably as long as they are not too tired, the choices do not require excessive calculation, they are not thinking about unfamiliar problems, etc. As with semantics and pragmatics, we can try to incorporate those caveats and qualifications into axioms and close, or develop open principles, whose application depends on context. Again, part of the answer depends on the goals of theory. I believe that broad maxims are more likely to reproduce the sort of open, speculative, hypothetical evaluation we make in our daily lives[54]. They also leave the baseline, rational setup untouched, somewhat like the literal meaning upon which we build[55].

Now this open-ended form of interpretation points to a theme that was raised both in 5.1. and in 5.3.1.2.: indeterminacy. It is intuitive that there often will be many plausible hypotheses that we may offer, each giving different theories about the agent, yet all equally compatible with all the evidence available. The point is rather general, as Quine (1964) shows with his famous thesis of the "indeterminacy of translation"[56]. As I noticed in chapter 2, sensory inputs are all the evidence we have to worldmake. It is also all the evidence we have to develop interpretations of other people's behavior, linguistic and other. Still, it is clear that we *always* go much beyond what is justified by these inputs. We offer conjectures, we project in this or that direction, we classify in ways that simply reproduce our own standards of similarity. All this in a path dependent, contextual and idiosyncratic form. There is thus no expectation that we can *reduce* our

[54] Cf. Tversky & Kahneman (1992, p. 317) "Theories of choice are at best approximate and incomplete. One reason for this pessimistic assessment is that choice is a constructive and contingent process. When faced with a complex problem, people employ a variety of heuristic procedures in order to simplify the representation and the evaluation of prospects. These procedures include computational shortcuts and editing operations, such as eliminating common components and discarding nonessential differences. The heuristics of choice do not readily lend themselves to formal analysis because their application depends on the formulation of the problem, the method of elicitation, and the context of choice."

[55] Incidentally, prospect theory has to deal with the problem of framing. notice connections with Schelling's work

[56] *Not* to be confused with the Duhem-Quine thesis of the underdetermination of theory by evidence. The indeterminacy of translation is additional to the underdetermination of theory by evidence. The underdetermination thesis tells us, in a nutshell, that several *theories* may equally fit all our evidence and still imply contradictory sentences, whereas the indeterminacy *further* tells us that there are many ways of translating one of these theories into the metatheory. For the relation between the two, see Quine (1970), for a critique of the Duhem-Quine thesis, see Quine (1975).

worldviews to sensory evidence, even if we accept that our worldviews are *caused* by sensory events. Quite the contrary: there are "irreducible leaps" in theory building[57]. Indeterminacy results: there are several theories that deal with all the *facts* equally well, and yet may be contradictory in some of the sentences they imply.[58]

Naturally, each of these possible theories about other people may have particular normative implications. Each will attribute more or less significant departures from rationality, and more or less error. Some portray other people in their best others in their worst *possible* light.

When someone commits herself to what sounds ludicrous, we always have the option of spending more time with her to try and understand what she could possibly be thinking. We may, of course, end up deciding that the person is wrong, and account for her error, or conclude that she is being devious in her interaction with us. But we often learn to see things in a novel way. Perhaps not a way that we find particularly interesting for any purpose, but coherent, reasonable, beautiful perhaps.

In any event, there *are* limits to how much error and incoherence we can attribute to another *person*. Indeed, there is a limit to how different others can seem before we stop counting them as people. What I am urging is that differences are, to an extent, on the eye of the interpreter. To count something as a person is to assume that it is reasonable, that it behaves in a way that is understandable by giving reasons. There are no hard and fast lines as to what should be counted as a person or as something else[59]. As Rorty

[57] Quine (1973, p. 178) urges that it is sound methodological practice to minimize these leaps: "Don't venture farther from sensory evidence than you need to. We abandoned radical empiricism when we abandoned the old hope of translating corporeal talk into sensory talk; but the relative variety still recommends itself. We recognize that between the globally learned observation sentences and the recognizably articulate talk of bodies there are irreducible leaps, but we can still be glad to minimize them, and to minimize such further leaps as may be required for further reaches of ontology."

[58] More succinctly: "there seem bound to be systematically very different choices, all of which do justice to all dispositions to verbal behavior on the part of all concerned." Quine (1969c, p. 54) Part of the thesis of the indeterminacy of translation involves the thesis of the inscrutability of reference: "the only difference between rabbits, undetached rabbit parts, and rabbit stages is in their individuation. If you take the total scattered portion of the spatiotemporal world that is made up of rabbits, and that which is made up of undetached rabbit parts, and that which is made up of rabbit stages, you come out with the same scattered portion of the world each of the three times. The only difference is in how you slice it. And how to slice it is what ostension or simple conditioning, however persistently repeated, cannot teach." Quine (1968, p. 32)

[59] Cf. Rorty (1972, p. 659): "I doubt that we can ever adumbrate general ways of answering questions like, Is it a conceptual framework very different from our own, or is it a mistake to think of it as a

(1972, p. 657) puts it, "I do not see what 'known in advance not to be a person' could mean when applied to the butterfly save that the butterfly doesn't seem human. But there is no particular reason to think that our remote ancestors or descendants would seem human right off the bat either." History, unfortunately, supplies many examples of people being treating by other people as inferior.

As Adam Smith ([1790] 1984, p. 27) notices, "if we consider all the different passions of human nature, we shall find that they are regarded as decent, or indecent, just in proportion as mankind are more or less disposed to sympathize with them." The same way I argued in chapter 2 that having a broad mind is no substitute for hard work, the acceptance of ethno, species and even egocentrism as inescapable is no excuse for cruelty. Quite the opposite: awareness turns it into a choice. We can avoid what is bad about our ethnocentrism, and be cosmopolitan in our attribution of personhood and human dignity. As Martha Nussbaum (2010, p. xvii) writes, "seeing the shape of a human being before us, we always have choices to make: will we impute full equal humanity to that shape, or something less?"

5.5. Summary

In this chapter I discussed approaches that try to derive theories of action from behavioral evidence. The difficulties the approaches surveyed have in common result from their attempt to reproduce the pattern that rationality makes by ever more sophisticated syntax. From axiom to axiom, using first-order logic, economists have been trying to get closed theories that fit the behavioral evidence. Whatever the interest such endeavors might have, they do *not* throw light on the way agents actually make sense of each other in particular contexts of interaction. I have thus proposed that instead of closed axioms we pay attention to broad maxims, applied case by case. I have also noticed that there are usually many theories that fit all the evidence we have, each with potentially unique normative implications.

language at all? Is it a person with utterly different organs, responses, and beliefs, with whom communication is thus forever impossible, or rather just a complexly behaving thing?"

6. A Return to the Socialist Calculation Debate

Having discussed the nature of economic explanation, its dependence on agents' attitudes, and how to determine the latter, in this final chapter I focus on the Socialist Calculation Debate, the midwife of the Austrian self-awareness, to broach the fundamental issues that were then at stake: the relation between theory and practice, the possibility of certain forms of economic organization, and the nature of economic reality. I shall try to explain how the results from the previous chapters enable us to better understand these issues.

In 6.1. I offer a historical overview of the Socialist Calculation Debate in the 1930s. I try to show that, right from the time of Pareto and Barone, economists seem to have believed in the existence, "out there", of the "fundamentals" and "the" economic problem of a community. In Pareto and Barone this was not very explicit nor very important, in that they could also argue for the universal validity of economics in other ways. But when the socialists of the 1930s tried to answer Mises' challenge as they understood it, this view on fundamentals was not only perpetuated but, at times, *identified* with the thesis that economics has universal applicability. Moreover, they all but ignored the epistemological problem of applying abstract theory to applied problems. I criticize the socialist position with recourse to an orthodox interpretation of Hayek's work.

In 6.2. I apply the lesson of chapter 4 - that economic classification and signification depend on the attitudes of agents - and also the view elaborated in chapter 5 that there is much choice in the determination of agents' attitudes in order to show that what categories and notions *denote* is determined contextually and holistically. Ditto for economic problems. I emphasize the Austrian view of the market as a *process*, as opposed to a mechanism. My reflections appeal to some of Hayek's intuitions on dispersed knowledge, which I explicate in part with the notion of aggregative group knowledge: if *I* know how to drive, and *you* know the map, then *we* know how to reach the destination. My point is that it might *socially* follow that *we* know this, or that *this*

machine becomes a means of production. Indeed, Hayek's 'unintended consequences' are these 'consequences'.

6.1. Historical Overview

6.1.1. The universal validity of theoretical economics: Pareto, Barone and the early Mises

By the turn of the last century, it was common to hear that the problems studied by political economists, for instance the problem of value, would have no place in the societies of the future. These problems, the indictment went on, were peculiar to the coetaneous capitalist social organization, and they should not trouble the socialist in whose ideal society, which the future was bound to bring, all economic problems would be replaced by technical ones (Cf. Pierson (1902, p. 42)).

Many economists, however, were together in believing not only that economic problems arise in any community, whatever its specific political or juridical organization, but also that theoretical economics is relevant to their study. Enrico Barone was among those economists. In his famous 1908 article, he tried to elucidate how collectivistic economies were to function by applying the concepts of the 'Mathematical School' (1908, p. 246), i.e., the mathematical theory of general equilibrium. Similarly, Vilfredo Pareto ([1909] 1966, p. 362 or VI.52ff) applied the general principles of economics to study economic equilibria in collectivist societies.

Both authors concluded that, *conceptually*, the economic problems in a decentralized, and those in a collectivist, society are identical. In other words, that at the level of fundamental theory there is little to distinguish one from the other. Barone concluded that "all the economic *categories* of the old regime must reappear" (p. 289, italics supplied). The prices or incomes as known in decentralized societies may disappear, but not their theoretical surrogates. In his turn, Pareto ([1909] 1966, p. 362 or VI.54) wrote

that “les prix, les intérêts nets des capitaux peuvent disparaître … comme entités réelles, mais ils demeureront comme entités comptables.”

Interestingly, both authors’ arguments assumed that there are fundamental productive constraints in any community, be they stocks of primary goods, or techniques of production, which *actually* exist, independently of the community’s present form of organization. Thus, Pareto ([1909] 1966, p. 363 or VI.58 italics supplied) wrote that “l’État collectiviste, mieux que la libre concurrence, semble pouvoir porter le point d’équilibre sur *la* ligne des tranformations complètes”; which would be an uninteresting remark unless the actual line of productive possibilities were the same under both organizations. Barone (1908, p. 274), in turn, not only wrote that “the system of the equations of the collectivist equilibrium is no other than that of the free competition”, but also spoke of “socialized resources” (e.g. 267) which are clearly treated as pre-existing (cf. p. 265).[1]

I should remark that, in spite of defending the general applicability of the concepts of economics, and even identifying *the* economic problem of a society independently of its concrete form of economic organization, these authors do not diminish the practical differences between forms of organization. Thus, Pareto ([1909] 1966, pp. 233-4 or III.217) writes that it is inconceivable that the economic problem of any society could *actually* be solved with recourse to equations, as it would require computational powers beyond those imaginable[2]. Barone, in turn, draws attention to the different dynamic effects that alternative forms of economic organization would have on the evolution of the fundamentals. He (1908, pp. 288-9) chides conventional socialists that find in the constant renovation of the productive fabric of capitalist society a wasteful phenomenon with no place in a rational society. He argues that in any socialist society that does not wish to ossify in the present, large-scale *scientific experimentation* (p. 287-8) would *also* have to take place, with concomitant “waste” of resources. In other words, Barone

[1] Similar assumptions of coherent identification of economic magnitudes across forms of economic organization are at work whenever we try to compare changes to an economy. When economists apply counterfactuals of the sort: “if competition were perfect in country Y, GDP *would be* x% higher than it is” it is obvious that in the counterfactual world it is assumed that the “fundamentals” are unchanged.
[2] Cf. Pareto ([1909] 1966, p. 234 or III.217): “si on pouvait vraiment connaître toutes ces équations, le seul moyen accessible aux forces humaines pour les résoudre, ce serait d’observer la solution pratique que donne le marché.”

argues that the opportunity cost of not wasting with experimentation is the foregone increased productivity of improved techniques of production[3].

Ludwig von Mises agrees with Barone and Pareto that economics could be used to study any human community. However, Mises defends his position not by appealing to the pre-existence of fundamentals, but rather by arguing that where there is action there is economizing. As he ([1920] 1990, p. 8) writes, "every man who, in the course of economic life, takes a choice between the satisfaction of one need as against another, *eo ipso* makes a judgment of value." According to Mises, we have an economic problem whenever to achieve an end, another end must be relinquished, i.e., when means are scarce *and*[4] have alternative uses. Whenever we choose one course of action as opposed to another, we need to judge and to *calculate* which one is best, and therefore the abstract study of choice is relevant wherever action takes place. As will be apparent, Mises' viewpoint has wide-ranging implications.

6.1.2. Mises's Impossibility Argument

Mises ([1920] 1990, pp. 8-9)'s definition of economic problem is broad, counting under its scope both the commonplace, whether to have chicken or salmon for dinner, and the complex, whether and how to build a new rail line between Paris and Vienna. Still, whereas the former most people can solve off-hand, solving the latter in any interesting sense requires a different kind of data collection, computation, speculation, and entrepreneurial judgment. Mises ([1920] 1990, p.13) submits that without an intellectual[5] division of labor, a human mind would stand perplexed "among the

[3] A similar point was later made by Mises ([1932] 1981, p. 103) and Hayek (1940, p. 199), too.
[4] Cf. Robbins (1935, p. 13) and Hayek (1935a, pp. 122-123).
[5] "The distribution among a number of individuals of administrative control over economic goods in a community of men who take part in the labor of producing them, and who are economically interested in them, entails a kind of intellectual division of labor [eine Art geistiger Arbeitsteilung], which would not be possible without some system of calculating production and without economy" Mises ([1920] 1990, p. 12). It should be noticed that instead of "intellectual" the adjective "mental" could also have been used, as indeed it is in the translation of Mises's *Gemeinwirtschaft* (Mises ([1932] 1981, p. 101)). In light of subsequent controversies over the differences between Hayek and Mises (e.g. Salerno (1990) (1994), Rothbard (1991), Yeager (1994) (1996)), the "mental" rendition might have prevented confusion. Hayek (1937, p. 50) actually uses division of *knowledge* when quoting Mises.

bewildering mass of intermediate products and potentialities of production" that were available in the society of his time. In fact, it is doubtful that concrete problems of more than household complexity could even be posed.

Mises argues that in an economy in which complex economic problems are constantly being solved, there must be some economically meaningful process of intellectual division of labor. He found such process in the price system resulting from "the interplay [Zusammenspiel] of the subjective valuations"(Mises ([1932] 1981, p. 103)) in concrete exchanges against money, across the productive structure of the economy. Being the cumulative outcome of numerous exchanges against a common medium, prices can be used by economic agents as summaries of relevant information when making economic judgments[6]. Prices are not perfect, of course: they are not a complete, flawless index of alternatives according to some ideal standard. Among other reasons, there are many choices that are not touched off by monetary exchange. Still, as long as the range of monetary exchange covers a significant part of the productive structure, we are told that the price system is "sufficient" (Mises ([1920] 1990, p. 11))[7].

But, moreover, Mises argues that such widespread monetary exchange is actually *necessary* for an economic, global coordination of all the actions that result from local evaluations. In other words, he defends that no advanced economy can be without a price system with sufficient scope. His argument for this seems to be that the economic significance of prices results from their being a precipitate of *competitive* market exchanges. And no other system of calculation could have the same economic significance as the price system without also being tied up to a vast array of "market dealings" (p.18).

[6] Mises' position has similarities with that of Nicolaas G. Pierson (1902). In contrast to Pareto and Barone's moneyless economies, where from pre-existing fundamentals (preferences, resources and coefficients of production), a solution to *the* economic problem is obtained arithmetically, Pierson emphasizes the incommensurability of goods when not reduced to a standard of value. As he writes, "we cannot subtract cotton, coal and the depreciation of machines from yarns and textiles" (p.70). This calculation requires "evaluation" (p.82), and is done through monetary exchange (p.72). "It is not a purely technical problem which is here in question, but rather a *decision* as to the most profitable way of employing material things; and the rightness of such a decision must depend upon the rightness of the *evaluation* which preceded it" p. 59 (italics supplied).

[7] Whether we could spell out the "significant" and the "sufficient" without circularity is an open question.

If we define a socialist economy as an economy with socialized means of production *and* without monetary exchange against the means of production, his famous impossibility[8] result follows from his premises[9].[10] To this result I now turn.

6.1.3. The Received Interpretation of Mises's Impossibility Argument in the 1930s

Mises' early challenge to socialism has been interpreted in different ways, and has spurred much discussion. The most famous debate which stems from his argument is the *Socialist Calculation Debate* of the 1930s. It is important, however, to separate our exegetical preferences, such as the interpretation in the previous section, and later theoretical developments inspired by Mises, from the received interpretation in the 1930s. If we already read too much in the early Mises, we may be uncharitable to the market socialists who attacked Mises' ideas[11].

---

[8] For Mises, the impracticality (Undurchführbarkeit) of socialism and its impossibility (Unmöglichkeit) are the same proposition. Mises's loose use of language in his early articles was explicitly noticed by Hayek (1935a, pp. 145-6). *Impossibility* is definitely closer to his meaning (Cf. Salerno (1990)).

[9] This definition of socialism was not a straw man. The important philosopher Otto Neurath (1919) discussed social organization with calculation *in natura* (Cf. Chaloupek (1990, p. 673)). Barone's (1908, p. 267) collectivist economy was also moneyless. The early Mises did not see himself as arguing against marginalists. That as late as 1932 he still saw all types of "subjectivists" as saying essentially the same thing under different garbs is shown by Kirzner (1988, pp. 9ff). For a survey of the German-language debate on socialism during the 1920s, see Chaloupek (1990). Keizer (1987) examines Mises's two interventions in this debate. Chaloupek argues that this debate was intellectually inferior to that in the 1930s, an assessment with which Mises ([1996] 2007, p. 706f) certainly agreed.

[10] In Mises ([1932] 1981, pp.186-7) and Mises ([1920] 1990, pp. 23ff), he also worries about the problems of "responsibility and initiative". As he writes, "communal enterprise [lacks] the spirit of the business man." It is not through training, remuneration schemes, or the "moral purification of mankind [die sittliche Läuterung der Menschen]" (Mises ([1932] 1981, p. 191)) that this difficulty is solved. Entrepreneurs cannot be trained; clerks can. Nothing short of fully-fledged capitalism can make the entrepreneur economically responsible for the profits and, more importantly, for the losses that result from her choices. Finally, he writes that these "problems (…) do not arise from the moral shortcomings of humanity. They are problems of the logic of will and action which must arise at all times and in all places" (Mises ([1932] 1981, p. 191)). This second line of attack was all but ignored by his socialist opponents in the 1930s on the grounds that capitalism was no less bureaucratic than socialism. See, for instance Lange (1937, pp. 123ff), who thinks that this was a problem for sociologists to deal with, not economists. Lerner (1937, p. 267f) agrees. See also Durbin (1936, p. 678). Interestingly, Lerner (1938, pp. 74-75) states that "to agree that managers will not manage prudently unless they manage with their own money is to agree with von Mises that socialism is impossible." This dismissive attitude was soon overcome, and problems with incentives became a central concern of the economics of planning. Cf. Bergson (1948)

[11] Salerno (1990), for instance, uses 'competition' seven times in a postscript to Mises (1920), and Lavoie ([1985] 2000, Ch. 3) uses the expression 'entrepreneurial rivalry' when summarizing Mises's position at

In a famous two-part article, Oskar Lange (1936, p. 54) accuses Mises of mixing up two senses of the word price[12]. According to Lange, in Mises price sometimes stands for exchange ratio, other times it stands for "terms on which alternatives are offered." If we are given a preference scale and a list of primary factors, such terms "are determined ultimately by the technical possibilities of transformation of one commodity into another, i.e. by the production functions."

According to Lange, if it is true that exchange ratios satisfactorily reproduce the fundamental terms, it does not follow that we cannot develop some other way of recovering that information that does not rely on competitive exchange. In other words, Lange does not find in Mises any argument that establishes more than the *sufficiency* of prices, and thus he remains unpersuaded of their necessity. Also, Lange (1936, p. 57) mentions that his arguments are not really using much that goes beyond Barone (1908).

This interpretation of Mises' challenge may be schematized as follows:

D1: Definition: An advanced economy is an economy in which complex economic problems are solved.

Proposition: An advanced economy is possible if and only if there is a way of coordinating decentralized action in an economically reasonable way.

P1: Sufficiency: The price mechanism resulting from the widespread interplay of subjective valuations against money across the productive structure of the economy is a way of achieving such coordination.

P2: Necessity: *Only if* there is a price mechanism resulting from the widespread interplay of subjective valuations against money across the productive structure of the economy can such coordination be achieved.

---

the time. Whatever their merit as reconstructions of Mises's argument, these positions owe much to hindsight.

[12] The distinction of these two senses is Wicksteed's (1910, p. 28): "'Price,' then, in the narrower sense of 'the money for which a material thing, a service, or a privilege can be obtained,' is simply a special case of 'price' in the wider sense of 'the terms on which alternatives are offered to us' (…) (how much of this against how much of that?)."

Conclusion: Without widespread monetary exchange, an advanced economy is impossible.

D2: Definition: Socialism is a form of advanced economic organization in which the means of production have been socialized and are outside the scope of monetary exchange.

Corollary: Socialism is impossible.

If we interpret Mises this way, it is clear that P2 and D2 are crucial for his impossibility argument. If we are persuaded by his case for P1, then we may say that he manages to prove the sufficiency of widespread monetary calculation for the articulation of the local with the global. However, whether we or the socialists are so persuaded is irrelevant to establish the main conclusion. Unless he persuades us that P2 is true, he has not established the necessity which an impossibility argument requires. If we agree that he has not proved P2, we may also make Lange's rhetorical point that there is merely a difference in emphasis from Barone and Pareto. Whereas the latter focused on the similarity of *the* economic problem of a society under a capitalist or a socialist organization, Mises focused on the difficulties in implementing a socialist society. But, as seen above, those two authors would not have denied such difficulties[13].

[13] Lange uses powerful rhetorical devices. He starts with his famous (and, undoubtedly, insulting) proposal of erecting a statue of Mises in "the great hall of the Ministry of Socialization" (Lange (1936, p. 53)). He then remarks that his own rebuttal is essentially that of Barone and Pareto who had answered Mises's challenge before he had even put it (p. 54). And he finishes by accusing Mises of the "Institutionalist" view that "economic laws have only historico-relative validity" (p.55). He thus starts with ridicule, moves on to question Mises's scholarship, and finishes by indicting his position as unscientific. Lange also spread the rumor that Hayek and Robbins had retreated to a second line of attack. According to this rumor, these two authors no longer found socialism impossible, but merely practically difficult (p. 56). Lange's positions are curiously echoed by Schumpeter ([1954] 1994, p. 989), who writes, for instance, that "it is not clear (…) whether von Mises really meant to deny the validity of the Pareto-Barone result. For it is quite possible to accept it and yet to hold that the socialist plan, owing to the administrative difficulties involved or for any other of a long list of reasons is 'practically unworkable'." For a discussion of Schumpeter's take on the Socialist Calculation Debate, see Moura (2014).

6.1.4. The trial-and-error procedure

As a matter of fact, Lange actually belied that P2 had already been proven false. A few years before, Fred M. Taylor (1929), in his presidential address to the American Economic Association, had expounded a non-tâtonnement procedure that, for Lange and other market socialists, could be used to construe a counterexample to P2[14].

According to this counterexample, we can solve the problem of imputation by trial and error. First, a list of the (basic), pre-existing primary factors is constructed by the Central Production Board, and tentative valuations (numbers) are given to each factor. Then, *the* productive units, managed by functionaries, choose the quantity of each product they are to produce by taking the prices of primary factors as given and following some predefined rule (depending on the specific author, it could be to equal price to marginal cost, or to average cost, to maximize profits, to minimize costs, etc.). Finally, citizens are allowed an income to be spent on products of their choice in a market for consumption goods. Depending on the valuation of consumers' as materialized in purchases, some goods will be in excess supply, others in excess demand. This is transmitted up the productive network by the mechanical rule-following of the functionaries, and results in surpluses and deficits of primary goods. New valuations consonant with these deviations are to be produced (again, by following

[14] I am using the expression 'non-tâtonnement' to mean that trade occurs at non equilibrium prices. See, e.g., Hahn & Negishi (1962). It should be remarked that Taylor himself had no ideological sympathies for socialism. On this, see Taylor (1925, Chs. XLII-XLVIII), and also Z. C. Dickinson (1960, p. 48): 'There are numerous evidences that Taylor did not regard his own demonstration [i.e. his trial and error procedure] as a statement of belief in the over-all [sic] feasibility and desirability of socialism."

prescribed rules), and the procedure goes on until equilibrium is established. Once the problem of imputation is solved, it is "easy" from there on.[15] [16]

It is important to point out that, as Karen Vaughn (1980, p. 537) notices, "the real debate (…) was among the socialists themselves." If, at first, their preoccupation was to meet Mises's impossibility charges, once they were persuaded that something along the lines of Taylor's procedure was sufficient for this, they turned to perfecting their market-socialist creation (cf. Bergson (1948)). Such attempts often led to important contributions to neoclassical price theory.

Take, for instance, the question of what rule the management of socialized firms is to follow. For Lange (1936, pp. 63ff), the price structure resulting from perfect competition has significance because it is parametric: unable to influence prices, agents take them as given, as parameters. The competitive price structure is the *only* one satisfying "the *objective* equilibrium conditions, i.e. [which] equalises demand and supply of each commodity." (p. 63, italics supplied). If the "parametric function of prices" is imposed by

[15] The key passage in Taylor (1929, p. 7 italics supplied) is: "(1) They would set about constructing factor-valuation tables in which they gave each factor that valuation, which, on the basis of much careful study, they believed to be the nearest approximation to its correct valuation which they could work out in advance of experience; (2) they would then proceed to carry on their functions as *managers* of all productive operations as if they considered the valuations given in their provisional tables to be the absolutely correct valuations; (3) while thus acting, they would after all keep a close watch for results which would indicate that some of their provisional valuations were incorrect; (4) if such results appeared, they would then make the needed corrections in the factor tables, lowering any valuations which had proved too high, raising any which had proved too low; (5) finally, they would repeat this procedure until no further evidence of divergence from the correct valuations was forthcoming." It is on page 8 that we learn that, once we become aware of discrepancies in the valuations, "the rest would be easy".

[16] H. D. Dickinson (1933) also develops a procedure based on "successive approximation" (p.231). His framework is similar to Taylor's for the first iterations. But once "the system has got going" (p.242), he tells us that we may dispense with the trial-and-error setup, and simply solve the problem mathematically, with recourse to a number of demand and production functions, the latter being essentially "a matter of technical rather than economic calculation" (p.242). Dickinson states that the problem is not even one of solving simultaneous equations but one of calculus of variations since "it is one of small deviations from an already established equilibrium" (p.242). After Robbins's ([1934] 2009, p. 151) and Hayek's (1935a, pp. 152ff) criticisms (and, I might add, Pareto's, as seen above) to the effect that solving the economic problem with recourse to equations is absurd, if for no other reason for computational difficulties, this technique was deemed implausible by the socialists and soon abandoned. Durbin (1936, pp. 677-678) approvingly quotes Hayek and Robbins on this, and Lerner (1937, p. 254) states that the adoption of such a technique stemmed from a confusion between the system "of analysis and the technique of administration." He writes that "except upon the basis of some such confusion it is difficult to imagine any economist putting forward as a technique of administration a scheme, like Mr. Dickinson's, in which all the equations have to be solved before any economic decision is reached." (p.254) Lerner is not being, strictly speaking, fair on Dickinson, since the equations would only fine-tune the path to the (*ex hypothesi*s stable) equilibrium.

the Central Production Board in the trial-and-error procedure, i.e., if agents are made to take *every* price as given, then the accounting prices resulting from the trial-and-error procedure will also reproduce the objective equilibrium conditions of the economy. Lange (p. 66 italics supplied) goes so far as to write that "the accounting prices in a socialist economy can be determined by the *same* process of trial and error by which prices on a competitive market are determined."

In fact, Lange (1937) believed that the trial-and-error procedure retains the virtues of competitive markets whilst getting rid of many of its vices. The problem of externalities, for instance (pp. 125ff), would be solved by making prices more comprehensive, and the business cycles would be kept in check by insulating local disturbances and thus preventing systemic effects (p. 126). With his characteristic irony, Lange argues that actual capitalism is much closer to what is found in the books of Chamberlin or Joan Robinson, than in the books of Marshall or Walras: "as a result, Professor Chamberlin and Mrs. Robinson face the danger of losing their jobs under socialism, unless they agree to be transferred to the department of economic history to provide students of history with the theoretical apparatus necessary to understand what will appear to a future generation as the craze and folly of a past epoch." (p. 127)

But, as Abba Lerner (1936) observes, perfect competition is not an end, but a means to the real end - the social optimum (p. 74) or "the most economic utilization of resources" (Lerner (1937, p. 256)). If, as Lange (1936, p. 67) suggested, managers of socialized firms make average cost equal price, the technological realities that lead, in the long-run, to the ubiquity of big firms under modern capitalism will make prices greater than marginal costs, in which case production may be *smaller* than optimal. But, with such cost-structures, if average cost equals price, then marginal cost is greater than marginal revenue, in which case production may be *greater* than optimal. Lerner (1937, p. 255) writes that "no proper choice can therefore be made between these two symptoms until we know their relationship to the true desideratum." He writes that "we must therefore aim *directly* at our real object" (p. 256 italics in the original). In other words, the optimal management of these firms cannot be solved by abstract theory alone. [17]

[17] An author who in the 1930s criticized *market* socialism was Maurice Dobb. According to Dobb (1939), market socialism would inherit "two of the principle vices of capitalism" (p.714), to wit, chronic

Hindsight now offers further difficulties, even abstracting from problems with implementation, and from dynamic issues[18]. The modern reader is struck by the strong assumptions behind the trial-and-error procedure. Why does an equilibrium exist? And why just the one? Even if it exists and is unique, why would a non-tâtonnement procedure converge to it? And would *this* specific procedure converge to it? What is meant by convergence to begin with? If it is taken in its mathematical sense, then under any given finite amount of time how close would we be to an equilibrium? Would we only get there with the Second Coming, meanwhile going through Great Leaps Forward and potato famines? Finally, if there is convergence to an equilibrium, would it be the best one? Is optimality a coherent notion in this context? Would the equilibrium be good in some sense that ensured any superiority with respect to the status quo? How would it depend on the initial distribution of income?

All these questions and many others were raised in the following decades[19] [20]. What is important to observe, however, is that even though they cast doubt on the potential practicality and even the theoretical cogency of this particular form of *market* socialism, and offer us requirements that any alternative along Walrasian lines would have to observe, these questions do not dispose of Lange's criticism of P2. This criticism results from Lange's distinction between exchange prices and the fundamentals whose terms

unemployment, and instability. As he (1940, p. 277) puts it elsewhere: "those who dream of marrying collectivism to economic anarchy must, at any rate, not pretend that the progeny of this strange match will inherit only the virtues of its ill-mated parents." Dobb emphasizes the importance of coordination, and he finds it inconceivable that "decentralized autonomy" (1939, p. 728) and the appeal to simple rules (e.g. price equal to marginal cost) could improve much over capitalism. He advocates *proper* central planning, based on the "fuller knowledge of the data" available to the central planning board, which is "a crucial element in the superiority of a socialist over a capitalist economy" (p.726).

[18] On the problem of setting up an investment policy under the market socialist procedure, see Lange (1936), Lerner (1934, pp. 72-73) and Dobb (1939).

[19] See Arrow & Hurwicz (1977, pp. 44-5). Geoffrey Heal (1973) offers a thorough account of the state of the art of the economics of planning in the early 1970s, incorporating insights from the many developments in mathematical economics since the 1930s. See also Heal (1986).

[20] An example of a model in Taylor's spirit that tries to deal with these issues is found in Arrow & Hurwicz (1960). They develop a tâtonnement process with three levels of decision: a custodian for each good, the managers of the several units of production, and a helmsman defining final demands (p. 69-70). Each level would have a particular problem to solve, and should function on as little information as possible: so managers should only know the technology of their units, and the prices set by custodians, the helmsman should only know the prices and the welfare function, and finally the custodian of each good should only know the aggregate excess demand or supply of its product. A gradient method is developed that converges to the social optimum if strict concavity is liberally assumed throughout. This article exemplifies the rather crippling assumptions that have to be made once we are (syntactically) realistic about informational limitations and the meaning of convergence, even in a purely Walrasian framework.

they are supposed to reproduce. *If* the fundamentals exist independently of the former, then it does not seem *prima facie* impossible that we determine the solution to the economic problem of our community without resorting to prices resulting from a competitive exchange of means of production, even if Taylor's trial and error procedure is unconvincing.

6.1.5. Abstract vs. Applied Problems

In 6.1.1. I contrasted the belief in the independent existence of fundamentals, such as stocks of primary goods, techniques of production or particular preferences, with the belief in, to use Lange's (1936, p. 55) expression, the "universal validity of the fundamental principles of economic theory". Using the example of Mises, I noticed that the arguments for the universal validity of economic theory need not support the existence of fundamentals independent of contingent economic organization.

But the applied theorist faces a further problem. Perhaps there is no reason to question the existence of universal economic notions, i.e. notions that appear in any theory that would interest us, and that have constitutive syntactical properties and interrelations which allow us to derive non-trivial, general conclusions. We may not even question, although in 6.2. I shall, that besides these universal *notions*, there are specific *entities* with an economic role, say that of being resources or techniques of production, that is independent of the contextual specificities of a community's practices. But, even assuming this, once we leave pure theory we must face the *further* problem of applying these abstract, general notions to specific circumstances, to solve, *inter alia*, the problem of determining what the fundamentals actually are.

Indeed, how do we go from abstract microeconomic theory to, say, a theory of the market for chemical compounds in Germany, or a theory of consumption in Odysseus' household? Applied theory needs applied axioms: axioms such as 'the coal found at such location, when used in such a way, leads to the profitable production of so many units of acetic acid', or 'Penelope prefers to wait for Odysseus than find a new husband.' In other words, we want to determine a denotation for economic designators

in concrete circumstances. What procedures we should follow and what evidence we need to collect to give a reasonable denotation to designators and to ascertain the truth of applied axioms are questions that may only have historico-relative answers, *even if* theory has universal validity *and* the role of some entities is independent of concrete economic practice.

The authors who propose the trial-and-error procedure are insouciant about these problems of applying theory. Lange (1936) has no qualms, for instance, in saying that 'the amount of resources available' is as much a given under capitalism as it is under socialism (p. 55). He (1936, p. 55) goes so far as to state that *since* Mises holds that "without private ownership of the means of production no determinate index of alternatives exists", *then*, for Mises, "the economic principles of choice between different alternatives are applicable only to a special institutional set-up." Only an "Institutionalist" would hold such a position, Lange claims. Similarly, Taylor (1929, p. 3) defines primary factors as (italics supplied) "those *economic* factors of production *behind which the economist does not attempt to go*, for example, the land itself; the water powers; the original raw materials such as metallic ores; the different kinds of labor services, *etc*."

But does this 'etc.' work from the perspective of the applied theorist of the *actual* socialist state in which there must be a real list of *all* the primary factors? Already in Barone (1908, p. 267 italics supplied), we are told that the "Ministry [of production] maintains, for no other purpose than the social accounts, *some method* of determining ratios of equivalence." But Barone does not justify the belief that there could actually be one such method, much less does he tell us what it would look like. As we get ahead in time, we find Taylor's 'etc.' repeated in Heal (1973, p. 7) who writes that 'resource constraints are more or less self-explanatory: the inputs of certain raw materials – labor, coal, land, etc.'

If this 'etc.' is, arguably, innocuous to the pure theorist, it hides a problem of applied theory that is a crucial problem to the socialist. This problem, as I will argue in the next section, Hayek shows to be an economic problem in itself.[21]

6.1.6. Hayekian Objections

It is not surprising that Hayek (1940, p. 188) writes that market socialist proposals have "been born out of an excessive preoccupation with problems of the pure theory of stationary equilibrium." Indeed, one of the best known objections of Hayek's against market socialists, expounded, *inter alia*, in Hayek (1945), is precisely that, even if it is legitimate for the economist, when doing abstract theory, to assume that she has available all the inputs that *would* be necessary to calculate solutions to economic problems (e.g. agents preferences, or the means and techniques of production), it is not usually legitimate to assume that there is any actual agent who does or even *could* command all the pieces of information necessary to calculate a solution when solving applied problems. To the contrary, the study of the acquisition, treatment and transmission of all sorts of information between economic agents is a crucial part of the problem of intersubjective economic coordination, and is relevant not only when we study how economic problems are actually being solved in the economy, but also when we wish to develop other procedures to somehow solve them. In both cases, we need to study who is informed of what, whence that information came, how they transformed information and other inputs into true and false beliefs, and how dispersed information and knowledge are *economically* coordinated and transmitted across the relevant social body. Whether it is *possible* that somebody ends up knowing all the facts necessary to solve applied economic problems, or that decentralized bodies transmit information and coordinate plans of action in an economically reasonable way, are not trivial questions.

[21] This insouciance about the determination of the relevant, available evidential bases, and the procedures for the obtainment, processing, and interpretation of evidence is by no means unique to theories of economic planning up to the 1990s (cf. Buchanan ([1969] 1999)). Many articles *identify* terms in theoretical models with concrete aggregates of data.

According to Hayek (1948b), economists' traditional view of competition "*assumes* the situation to exist which a true explanation ought to account for" (p.94, italics in the original). Perfect competition requires homogeneity of the commodity being traded, a sufficient number of buyers and sellers to deprive any of them of price-making powers, and complete knowledge of all the relevant factors (p.95). But if everybody knows what should be produced and what is the best way of producing it, what is competition about? These are precisely the questions that the *process* of competition is to answer. "It is only through the process of competition that the facts will be discovered" (p.96). For Hayek, "'perfect' competition means the absence of all competitive activities" (p.96).

In a world of constant change, of dispersed information and historical accident, "the solution of the economic problem of society is … always a voyage of exploration into the unknown, an attempt to discover new ways of doing things better than they have been done before" (p. 101). Competition is a dynamic, social process that coordinates the "dispersed bits of incomplete and frequently contradictory knowledge (sic) which all the separate individuals possess" (Hayek (1945, p. 77); see also p. 85).

One possible interpretation of Hayek's challenge to socialism is to accept that knowledge of the facts necessary to solve the economic problem of a community is somehow dispersed in people's heads, implicit in their routines, behavior etc., and that we may not be able to make such dispersed knowledge either propositional or succinct enough for it to be transmitted, or get the right incentive structure for it to be revealed, and therefore for it to be incorporated into any explicit mechanism of economic management. This interpretation of Hayek's objection became increasingly popular as economists concerned themselves with the informational properties and incentive structures of social systems.[22]

[22] By the 1970s these issues were at the center of the literature on the economics of planning, but, much earlier, Bergson (1948, pp. 216ff), in a well-known survey article, considered Hayek's objections important, especially with respect to "the problem of administration". Many later developments in economic theory can also be traced back to the influence of Hayek (1945). The Nobel Prize laureate Leonid Hurwicz was directly influenced by this and other articles of Hayek's, and dedicated an important part of his career to studying the decentralization of information, and informational constraints as feasibility constraints to several conceivable systems of economic organization (see, e.g. Hurwicz (1969)). More generally, mechanism design and information economics have also been said to have sprung out of this work of Hayek's (cf. Stiglitz (1996)). It is thus not surprising that Hayek (1945) should have been chosen in 2011 as one of the 20 most important articles published in the *American Economic Review* during its first centenary (cf. Arrow et al. (2011)).

This line of work clearly shows that the acquisition and transmission of knowledge of the relevant facts is an economic problem in itself. Hence, whatever the type of procedure we follow in order to determine the denotation of economic designators, we will have to face problems of acquisition and transmission of information that cannot be divorced from concrete economic practice. Hayek's criticism to the trial-and-error procedure, according to this interpretation, is that it assumes that the relevant facts are already known or, perhaps, that they could be easily determined. Hayek argues that the discovery of the relevant facts comes from the very process of solving economic problems, and the way knowledge and information are obtained and transmitted by competitive market exchange leads to the coordination of agents' decision-making. It is not clear what procedure the socialist could come up with that would have similar properties.

Even though this is a powerful, some might say definitive, objection to the trial-and-error procedure described in the 1930s, it is not yet clear that some improved form of decentralized market socialism could not skirt it. The socialist may agree that there are these difficulties. But in a market socialist setup there are decentralized agents and production units with independent management. Why could the latter not work as well as most modern corporations? And if we raise the objection that most of what agents know is discovered through *competitive* market dealings, the socialist could answer that it is always possible to give some leeway for new enterprises and market testing. What sort of crippling assumptions would have to be assumed to deal with matters of implementation, and incentive constraints can easily be imagined. In fact, if past experience repeats itself, the likely outcome of such an exercise is that the socialist will realize she has to grant more ground to the Hayekian, blurring the difference between market socialism, and decentralized market process *tout court*.

Naturally, after shifting ever more ground, the most interesting issue may well be "what of 'socialism' is left?" Indeed, some (e.g. Hayek (1935b, p. 177) or Hayek (1940, p. 186)) find market socialism not so much a criticism of what in 6.1.3. I named P2, but a replacement of D2 with something else, and therefore the abandonment of any *economically* interesting notion of socialism. Market socialism could perhaps be called socialism in a *juridical* sense, since "property" in the means of production (whatever

they are) is held by 'the State'. But is it economically a form of socialism too? As Roberts (1971, p. 564) put it, no socialist ever "believed in, placed hopes in, or fought for" market socialism[23].

6.2. On the determination of fundamentals

In 6.1. I discussed the historical episode that gave Austrian economists their identity. The debate enables us to distinguish three propositions that sometimes appear mixed up: that economics has universal validity; that the economic role of entities is independent of contingent social processes; and, finally, that it is possible to acquire and integrate the necessary pieces of information to solve the economic problems of a community without recurring to competitive market practices.

In the last two subsections, I explained that even assuming the universal validity of economic theory, *and* the independence of entities' economic roles, Hayek has shown us that the trial-and-error procedure faces grave shortcomings by ignoring the economic problems of knowledge. I interpreted his work as showing that the discovery of economic facts is, itself, an economic problem. The matter remaining, that of the independent existence of fundamentals, is the subject of 6.2.1. In 6.2.2. I employ the lessons of chapter 4 and also of chapter 5 in offering a preliminary reflection on the working of the market process.

[23] I should also add, though each side of the debate held a number of preconceptions that were foreign to the other, the most serious preconceptions were not ideological. To this I can attach Hayek's (1945, p. 89) authority. As he points out, "the differences can indeed no longer be ascribed to political prejudice. The remaining dissent seems clearly to be due to purely intellectual, and more particularly methodological, differences." That the socialists were willing to review their positions is illustrated by Lerner (1944) who owns that he had "overemphasize[d] its [socialism's] importance" (p. viii) and that "the fundamental aim of socialism is not the abolition of private property but the extension of democracy" (p. 1); and by Bergson (1967), who writes that "Hayek argued that such a result [proper managerial incentives] might not be easy to achieve. … the construction of a satisfactory incentive system now appears *more difficult than I envisaged it to be previously*" (p. 658, italics supplied). Schumpeter (2010, Ch. 16) is an example, on the other hand, of someone not sympathetic to socialism but who was instrumental in the acceptance of what Lavoie (1981) calls the 'Standard Account of the Debate', i.e. that the Austrians had lost. Frank Knight, as Boettke & Vaughn (2002) argue, was also, ideologically, clearly on the Austrian side, but rejected the idea that economics gives theoretical foundations to criticize socialism.

6.2.1. The fundamentals as socially contingent

In 6.1.5. and 6.1.6. I criticized the belief that the denotation of economic designators can be trivially determined. I argued that the epistemological problem of finding out economic facts is, in itself, a difficult economic problem, one of the many economic problems that are constantly being posed and given solutions as social processes unfold.

But we have also observed several authors speaking of the fundamental economic magnitudes, such as means and techniques of production, as if they could be determined outside of concrete social interaction, and, indeed, as if they *existed* independently of such interaction; in other words, it is as if something were, economically, a means of production, consumption good, etc. by way of an asocial set of properties that is prior to, and presupposed by, its economic appropriation by some agency. Taylor (1929, p. 3), for instance, when writing that primary factors are those "behind which the economist does not attempt to go", implies that their being primary factors is a non-economic fact, non-economically determined, and thus a given to the economist. Taylor seems to hold that they may be correctly described as economically relevant primary factors whatever the specific form of economic organization. Economic organization is thus, ontologically, a mere contingency.

According to this view, means of production are given by the engineering department, techniques of production by the book of blueprints, consumption goods by a menu. The a-sociality of the fundamentals permits not only their ascertainment outside of social processes, but their identification across forms of economic organization: to speak of *the* problem of one community, and to compare the several solutions given to it under radically different arrangements. It might not be too speculative to say that the development of circumscribed, often formalized, theories has led to the belief that the seemingly transparent, well individuated and strictly interrelated entities presupposed by those simple theories are also transparent, well individuated and strictly interrelated when more complex theories are developed.

In truth, it is often the case that a theory presupposes entities that are, *in that theory*, independent of one another, but that would not be presupposed or be independent were the theory incorporated into more general reflections about the economy. For instance, just because a simple microeconomic theory presupposes that there is a set of preference relations over a set of goods that is independent of what happens in the markets for those goods, it would be foolish to expect that, in a more complex theory, it would still make sense to believe that preferences were so independent, or even that markets for *such* goods existed.

One probable source of confusion is homonymy: because something is called a means of production in a theory of engineering, we assume that the same *thing* is, economically, a means of production. We even assume that it is an entity in the ontological domain of economics. If it is undeniably true (cf. 4.7.) that there are interrelations between the notion of means of production of the engineer and that of the economist, we have seen in 4.5. that not only are they not identical notions, the designators involving them do not, in general, refer to the same entities. There, I showed that the economic role of a pre-individuated entity may change without there being any change in the entity itself. It is not because some entity has certain physical properties that it is a means of production, a consumption good, or even a technique in an economically interesting sense. Physical specificities are not irrelevant, of course, but economic *function* depends on its place in an autonomous, holistic web of significance.

In an interesting passage, Schumpeter ([1954] 1994, p. 989 italics in the original) writes that "there is also a purely theoretical anti-socialist argument (sponsored by Professors von Mises, von Hayek, and Robbins), which is definitely wrong, namely the argument that, although there *exists* a determined set of solutions of the equations that describe the statics of a socialist commonwealth, there is, without private property in means of production, no mechanism by which to realize them. They can be realized by the method of 'trial and error'." I think that section 6.1.6. is sufficient to dispose of the assertion that the trial and error procedure is sufficient to realize the solution. What I now wish to question is rather the coherence of saying that there *exists* a determined set of solutions.

To a mechanistic view of the economy, determined by fundamentals which we may know independently of the very social processes they underlie, I opposed in 6.1.6. a Hayekian view where knowledge cannot be obtained outside of such processes. I am now pressing that the fundamental facts are, themselves, a precipitate of the economic process. In particular, that whatever role an entity has, or whatever entities take this or that role, depends on contingent social processes.[24] In other words, my reflections point to the conclusion that the same process that *discovers* the relevant facts also *creates* those facts. In the next subsection I address the *working* of the market process.

6.2.2. Economic creativity and group knowledge

Although I have not found much fault with the thesis, properly understood, that economic theory has universal validity, it should be recalled that what we find interesting, the ways we classify stimuli, and even the paths of association we follow are, as argued in chapter 2, the result of our history, of the language we use, and of the practices adopted in our communities. Economic science is done by people in particular social and institutional contexts, and therefore is, itself, an effect of social processes. But being an effect of social processes does not prevent it from also being causally efficacious: I believe economics can and does make much difference to those processes, not just by influencing policy, but also by influencing the *ethos*, the way agents think about economic and political reality.

We could, of course, focus on how the economist is a voice of class interests, and economics helps perpetuate relations of power; or, instead, focus on how the economist tries to produce true theories, relevant to her community, and economics makes enlightened social activism possible. To think that the tension between descriptions like these makes them incompatible or contradictory is to miss an important theme of

[24] Cf. Marx ([1867] 2004, Vol. I Ch. 7, p. 289) "Hence we see, that whether a use-value [Gebrauchswert] is to be regarded as raw material, as instrument of labour or as product is determined entirely by its specific function [Funktion] in the labour-process, by the position [Stelle] it occupies there: as its position changes, so do its determining characteristics." And also: "Capital is not a thing, but a social relation between persons which is mediated through things" (Vol. I Ch. 33, p. 932).

chapter 2, and of section 4.7.. We have different vocabularies available which are used to play different games. What descriptions we emphasize, and what theories we should use, depend on our interests and goals.

In subsection 6.1.6., I showed how Hayek argues that through decentralized competitive processes agents discover all sorts of facts, and information is transmitted and transformed into knowledge. The knowledge and information integrated in such competitive processes far transcend any individual's ken. Similarly, when I presented Mises impossibility argument in section 6.1.2., I showed how he emphasizes the need for an intellectual division of labor. In his early work he says that a human mind would be perplexed by all the potentialities and actualities of a modern economy, and in his late work (e.g. Mises ([1996] 2007, p. 707)) he underlines the importance of the multitude of individual centers of action for open-ended, entrepreneurial appraisement that such division brings[25].

Both authors thus find in decentralized competitive processes the ability to integrate several viewpoints, and to coordinate plans of action into satisfactory solutions to the concrete economic problems that constantly arise and, as argued in the previous subsection, are a precipitate of the economic process. Such a perspective could be opposed to another, of the economy as a mechanism, there being an optimal solution to pre-existing, almost trans-social, economic problems[26]. But what can we say about such decentralized processes that is constructive?

[25] In his later work, Mises argues that agents must look ahead and appraise the future, i.e., speculate, before acting. An agent can be in full possession of what are, for him, all the relevant *economic* facts about the present and past, but, since uncertainty characterizes all human action, the agent must exercise *judgment*, he must speculate, he must appraise. For the mature Mises ([1996] 2007), since the world is uncertain and constantly changing, "action is always speculation", and thus "in any real and living economy every actor is always an entrepreneur and speculator" (p. 252). In his entrepreneurial capacity, the actor *appraises* future uncertain conditions from his vantage point, and acts always in the expectation of thereby removing (potential) dissatisfaction. In his "restlessness" and "eagerness" (p. 255) to make a profit, he will "bid higher prices for some factors of production and lower the prices of other[s]" (p. 336) as soon as he "*understands* that there is a discrepancy between what is done and what could be done." (p. 336, italics supplied).

[26] Naturally, a perspective such as the one I attribute to Mises presupposes what may be called a second-order concept of optimality: there is a comparison between forms of organization. When Mises says that socialism is impossible, he means that without widespread monetary exchange complex economic problems could not be posed, let alone be given reasonable solutions (as opposed to being *rightly* solved). Why does this matter? I wish to say that it matters not because a simple economy is worse according to some a-historical standard, but because *we*, who live in a complex economy, rightly or wrongly, probably do not want to live in a simple household economy. Naturally, our preferences, even our sense of identity,

Sometimes, Hayek writes as though decentralized competitive processes did the same that a mind would do if it possessed all the information dispersed across society. For instance, in Hayek (1945, p. 86), he writes that "the mere fact that there is one price for any commodity … brings about the solution which (it is just *conceptually possible*) might have been arrived at by one single mind possessing all the information which is in fact dispersed among all the people involved in the process."[27]

Now, when we talk about collective knowledge, there are many plausible things that we could mean. Rescher (2005, pp. 31-32) distinguishes between three ways a group of people may know something[28]. When we say that a group G of people knows a proposition p, we may mean that *someone* in the group knows that p, or that *everyone* in the group knows that p, or that p *follows from* other propositions that are known by elements of the group. To the first type of group knowledge Rescher calls distributive, to the second collective, and to the third aggregative[29]. The notions of distributive and collective knowledge present no difficulties. The notion of aggregative knowledge, on the other hand, is less transparent. It is also the most interesting of the three.

Simple, trivial, examples of propositions known aggregatively that are neither distributively nor collectively known are easy to devise. For instance, Jane may know that she will catch a train home if and only if one train stops at her station within the

---

are a function of the society in which we live, but none the worse for that: we cannot be expected to transcend our viewpoint. See also Hayek (2002, p. 10), who writes that "all that can be empirically verified is that societies making use of competition for this purpose [discovery of relevant circumstances] realize this outcome to a greater extent than do others – a question which, it seems to me, the history of civilization answers emphatically in the affirmative." Indeed, for Hayek, the relevant welfare criterion (cf. Kirzner (1973, Ch. 6)) is not an approximation to an allocation judged optimal by an omniscient being, but the "abstract order" (p.15) that makes individual "rational, successful action" (p.14) possible, as empirically corroborated in the coordination of plans.

[27] Interestingly, it was at about the same time that Hayek (1943b, p. 121f) said that "the way in which individual minds interact may reveal to us a structure which operates in some respects similarly to the individual mind." I should also point out that the thought experiment *assumes* that, given all the information, there would be something determinate for the central mind to *do*.

[28] He actually distinguishes a fourth notion of group knowledge that he calls expert knowledge. "A group knows something by way of expert knowledge if this is known to the general run of its experts" Rescher (2005, p. 32). This notion is an important particular case of distributive knowledge, but not relevant for us.

[29] Formally, a group G knows a proposition p distributively $K_d Gp$ iff $(\exists x \in G \ \& Kxp)$, collectively $K_c Gp$ iff $(\forall x)(x \in G \supset Kxp)$, aggregatively $K_a Gp$ iff $(\exists p_1)(\exists p_2) \dots (\exists p_n)(\exists x_1 \in G)(\exists x_1 \in G) \dots (\exists x_1 \in G)(Kx_1p_1 \& Kx_2p_2 \& \dots \& Kx_n \& (p_n \& p_1 \& p_2 \& \dots \& p_n \vdash p))$. The symbol $\vdash$ is Frege's turnstile, and in this context is telling us that p is provable from the conjunction of the p_i.These formalizations are Rescher's (2005;31-2)

next five minutes ($p \leftrightarrow t$), but not know whether there will be a train in the next five minutes (t). But the conductor of the next train heading to Jane's station knows that there will *not* be a train stopping in the next five minutes: it is running late (~t). If we bring their knowledge together ($p \leftrightarrow t$ &~t) it *logically* follows something that neither knows: that Jane will not catch a train home (~p). This is an example of a group of two people, and of a true proposition that the group knows aggregatively but that is not known by anyone in the group.

Naturally, we can raise the objection that 'aggregative knowledge' is a misnomer. Whereas in the distributive and collective senses the group only knows what at least some of its members know, here we are saying that a group may know something that nobody in the group knows, which runs against our common-sense intuition that only individuals possess knowledge. Sure, the objector might argue, it may be a logical implication of the conductor's and Jane's knowledge that she will not catch the train, but since no one knows this, it is not knowledge at all. It is just a figment of *our* metatheory that incorporates what those *two* people know. It is no one's theory and therefore it is not relevant for anything anyone might want to do: it changes no one's actions, therefore it makes no difference.

There is truth to this objection. Still, the notion of aggregative knowledge just expands to the group the intuition that a person may not know every logical implication of what she knows. Similarly, a group of people may not know all the logical implications of what it distributively or collectively knows. What I wish to notice is the truth that *if* Jane managed to contact the conductor, she might have decided not to wait those five minutes and go straight home. *If only there had been a way* of transmitting the conductor's knowledge to Jane… There are opportunities here waiting to be seized, opportunities based on the exploration of the, in this case logical, interconnections of beliefs dispersedly held. The same way that an individual may, with effort, derive conclusions from held beliefs, social processes may turn aggregative knowledge into distributive or even collective knowledge.

But, I believe, the lessons are more general than matters of logical implication. There are *entities* whose existence may aggregatively, or *socially* follow from what distributively occurs in centers of *action*, the individual in the Austrian tradition. I believe that much of

what Hayek means by the study of 'unintended consequences' is, indeed, the study of what socially, or aggregatively follows from dispersed economic *events*. Also, the same way we can make sense of 'group conclusions' that go beyond what the elements distributively or collectively know, by analogy, the reduction of what socially follows to specific individual actions is to be expected only if there were some theoretically understandable uniqueness of what 'socially follows' from individual actions. Below, I will argue that there should be no such expectation.

Before I do so, I must warn against the peril of being blinded by the simplicity of examples. Naturally, when we take into consideration Jane's context at the train station, it is very easy to see what it would take for her to not waste her time waiting, and thus we can say that there *is* an opportunity that no one has yet seen. And we often can, without difficulty, analyse particular phenomena from their outside. It is from these intuitive possibilities that, I believe, error often comes. Indeed, the same transcendence makes no sense when we speak of the overall problem of society, which must include *the very thinking of the problem itself.* We can, of course, under many circumstances, leave an economic description and adopt another, perhaps common-sense view of matters. As shown in 4.5., if there are no necessary or sufficient structural properties for something to be, say, money, we can nevertheless correctly identify what items have the economic function of money in our community just by looking at them. But this rising above economic interrelations is *not* reasonable when we are speaking of overall economic organization, or, I believe, when discussing matters of policy with (expected) widespread implications. After a while the functional and the structural, as also shown in 4.5., part ways[30].

What people think and believe is often not logically related to the mental event that causes such thoughts and beliefs. There is no logical connection between an entrepreneur's noticing that his child needs a new pair of shoes and his sudden coming to have an idea for a business venture, yet one, for more or less obscure reasons, may

[30] As Mises ([1996] 2007, p. 392) writes, "it would be absurd to look upon a definite price as if it were an isolated object in itself. A price is expressive of the position which acting men attach to a thing under the present state of their efforts to remove uneasiness. It does not indicate a relationship to something unchanging, but merely the instantaneous position in a kaleidoscopically changing assemblage".

very well have caused the other. The patterns of association that have economic relevance are not reduced to logic and often are not easily articulated at all.

Moreover, there is also, as I showed in chapter 5, considerable choice in the ways we could and do interpret and make sense of others: several contradictory ways of making sense of other people are equally compatible with all the evidence we have and *could* have. We thus choose an alternative: to make sense of Jane as more or less rational, more or less foolish, as having this belief but not that, or vice-versa. Yet different choices may well have different economic effects.

Indeed, there is nothing determinate about our idiosyncratic processes of thought, be they our associations or our interpretations. They depend, *inter alia,* on context and personal history, and could often have turned out differently without there being any change in the economic dimension of affairs that could account for such difference. There is, from an economic perspective at least, *openness*. How plans of action coordinate, and the effects of their coordination or lack thereof on people's cognitive adjustments is, for all these reasons, quite indeterminate in the abstract. Indeterminacy and openness should not, however, be taken to be impediments to economic theory, but only to the interest of theory of a certain kind. We can, and should, study *how* social processes operate, and economic problems *are* solved in a reality *with such features*.

In his late work, Hayek emphasized ever more the distinction between simple and complex problems, and the difficulties that complexity brings. He (1963, p. 263) writes that "we know the general character of the self-regulating forces of the economy and the general conditions in which these forces will function or not function, but we do not know all the particular circumstances to which they bring about an adaptation. This is impossible because of the general interdependence of all parts of the economic process, that is because, in order to interfere successfully on any point, we would have to know all the details of the whole economy, not only of our own country but of the whole world." I am somewhat more radical than he is in this passage. He thinks it *false* that we could stand outside and wield such information, I think it is incoherent.

6.3. Summary

In this chapter, I offered an overview of the Socialist Calculation Debate. I studied its precursors, Pareto and Barone, Mises' famous impossibility argument, and also the proposals and answer to the challenge given by market socialists in the 1930s. In so doing, I distinguished three propositions: the universality of economic theory, the a-sociality of economic facts, and the dispensability of competitive, decentralized processes to ascertain those facts. The three propositions do not imply one another.

I concluded that theory, facts, and knowledge of the facts emerge *together*. The outlook of economics is situated, and relative to the social processes from which it derives, but, when looked from within, its outlook is as universal as it gets. Yet, even this economics cannot support the a-sociality of facts, nor does it imply the possibility of their independent determinability.

7. Conclusion

Most of what we do is caused by coming to have beliefs that involve expectations about the behavior of other people. Sometimes we know some of these people, but more often we do not. When we go to the market, for instance, we tacitly expect a blurry group of people to have acted in such a way that a complex process of production and distribution of goods of all orders cumulates in our finding the carton of skim milk we are looking for. Similarly for most of our decisions. Even when we make an explicit arrangement with a friend to meet for lunch, for our plans to come to fruition we tacitly rely on the appropriate behavior of countless strangers: that there is a bus on time, properly maintained, and with a knowledgeable driver, that drivers in general observe the rules of traffic, that people around us react in expected ways to our words and behavior, that the restaurant is open, with a functioning kitchen, waiters, etc.. And these are merely our first-order expectations: for the bus to be on time and properly maintained, countless other people must have behaved within a narrow set of possibilities.

Yet, most plans of action are tacitly made by people with all sorts of personal histories, worldviews, and contexts. Why and how do all these decentralized actions so coordinate that most of us can live fulfilling lives? This is the economist's oldest and proudest question.

I believe that the very formulation of the question betrays its subjective and intersubjective character. The economist is not concerned with movements of limbs, but with action, with plans, and these involve agents' intentions, and have meaning and significance to them. Hayek not only showed the necessity of understanding agents in order to explain their actions in an interesting sense, he was also aware of the methodological and logical issues that working with the vocabulary of the mental and the intentional involves. I thus dedicated chapter 3 and 4 to an exploration of Hayek's reflections on the moral sciences. I mostly worked with what I believe to be Hayek's most fruitful piece, his *Scientism and the Study of Society.* In this work Hayek tries to

articulate a slew of intricate thoughts. He is opaque and misuses words, he says impossible and even contradictory things. But he manages to address difficult issues in profound ways. As I tried to argue, not only did he understand the importance of intention and reasons for acting, he points in the right direction when he emphasizes the functional nature of economic concepts, and the particularities of the logic of the propositional attitudes. He also tried, even though in a very loose way, to explain *how* we could understand others. Since the moral scientist is a human, she shares a lot with her subjects. Her humanity is her entering point. Alas, he was not very clear, much less constructive about this.

Yet, the moral scientist needs to understand agents. To pursue the example I introduced above, if the bus does not come, we will try to find a cab, if people do not understand us, we rephrase, if the restaurant is closed, we grab a burger. All these other solutions may be more or less appealing but allow us to, in a sense, achieve our goal: to enjoy ourselves with our friend. I said something about these cognitive processes in chapter 2, where I discussed learning and inductive habits. But the moral scientist's main problem is, more generally, to learn how people make sense of each other, and how we could determine their worldviews and the meaning they attach to what they do. We need to be more constructive than Hayek was in the *Scientism* essay. To this end, in chapter 5 I critically explored revealed preference and decision theories from a Hayekian standpoint. The authors in these traditions have tried to recover agents' reasons for acting making explicit the behavioral evidence they condoned. I tried to assess whether we could use their proposals to explain the way agents, and moral scientists, *do* make sense of each other. Even though the answer is in the negative, the rubble helped me articulate important lessons. I argued that there is considerable indeterminacy in the possible theories that make sense of other people, each theory having different normative implications. How charitable to others and how cosmopolitan in our attributions of personhood are choices which people, and moral scientists, need to make.

Although I emphasized those two orthodox theories, there is a lineage of Austrian scholars, people like Peter Boettke, Ludwig Lachmann, Don Lavoie, Gary Maddison, David Prychitko, and Virgil Storr, who have taken a different route to reach an

understanding of agents.[1] These economists explore the radical subjectivist overtones in the Austrian tradition, and have been influenced by the hermeneutics of Wilhelm Dilthey, Hans-Georg Gadamer, Paul Ricoeur, and Alfred Schütz, by the work of Michael Polanyi and Max Weber, as well as by the anthropology of Clifford Geertz. For these Austrians the moral scientist should aim at producing "an interpretation of the way a people lives which is neither imprisoned within their mental horizons, an ethnology of witchcraft as written by the witch, nor systematically deaf to the distinctive tonalities of their existence, an ethnology of witchcraft as written by a geometer" (Geertz (1974, p. 29)), the dimension is that of the case study, and the goal is wisdom, not certainty. According to Lavoie (1991a, p. 6), "understanding is a 'fusion of horizons' between the author and his or her reader, where what is understood is not simply copied but is integrated with what the reader already knows, and in the process changed." To reach her goals, the economist may need to immerse herself in the communities she is studying, "mining their archives, reading their literature, listening to their folklore and praise songs, conducting interviews, and living amongst them" (Storr (2010, p. 37)). The thin descriptions used, namely abstract concepts like punishment, resource, etc. must be enriched by thick descriptions of the particular social context what is being studied. As Lavoie & Storr (2011, p. 221) put it, "it is as impossible to develop useful theory without an appreciation of history as it is to understand history without the application of theory".

Yet, the subjective is only half, if not merely one quarter of the economist's question. The intersubjective is missing. As Boettke (1995, pp. 28-29) asks, "how do social institutions bridge the seemingly unbridgeable gap between individual constructions of social reality and the regularity present in social order?" In chapter 6 I explored the Socialist Calculation Debate and noticed that market socialists in the thirties believed in the a-social existence of fundamentals like means and techniques of production, and that we could determine these fundamentals outside of the economic process. Their position involves a particular understanding of our question, and their trial-and-error procedure shows what they would count as an answer. I used Hayek to criticize their proposals and

[1] For an overview of this literature, see the articles in Boettke (2010), Prychitko (1995) and Lavoie (1991b).

tried to offer some of my own, ending with an analogy between what logically follows from premises, and what socially follows from decentralized action.

In his work after the 1950s, Hayek pays more attention to what he ([1982] 2012) calls "rules of just conduct on which the order that he [the economist] studies rests" (p. 4). As he (p. 105) writes, "the importance of rules of just conduct is due to the fact that the observance of these values leads to the formation of certain complex factual structures, and that in this sense important facts are dependent on the prevalence of values which are not held because of an awareness of these factual consequences"[2] Tony Lawson (1997, pp. 134-5) and Steve Fleetwood (1995) argue that it is in Hayek's late work that he "successfully integrates the notion of social structures into his analysis, thereby allowing the development of a notion whereby human agents navigate their way in the socio-economic world by following social rules of conduct." (Fleetwood (1995, p. 5))

The enduring, transfactual, operation of social structures such as Hayek's rules of just conduct gives the "material conditions" (Lawson (1997, p. 31)) for intentional action to be possible. Yet, much is still left open. As the example above shows, we often change our minds or readjust our plans in reasonable ways on the spur of the moment. All this is rather open in the sense that what happens could always have gone some other way.

[2] See also Hayek ([1982] 2012, p. 12) "we live in a society in which we can successfully orientate ourselves (…) because they [our fellows] are also confined by rules whose purpose or origin we often do not know and of whose very existence we are often not aware"; and Hayek ([1960] 2011, p. 230): "we can produce the conditions for the formation of an order in society, but we cannot arrange the manner in which its elements will order themselves under appropriate conditions. In this sense the task of the lawgiver I not to set up a particular order but merely to create conditions in which an orderly arrangement can establish and ever renew itself."

8. References

Allais, M. (1953). Le Comportement De L'homme Rationnel Devant Le Risque: Critique Des Postulats Et Axiomes De L'ecole Americaine. *Econometrica, 21*(4), 503-546.

Arrow, K. J., Bernheim, B. D., Feldstein, M. S., McFadden, D. L., Poterba, J. M., & Solow, R. M. (2011). 100 Years of the "American Economic Review": The Top 20 Articles. *The American Economic Review, 101*(1), 1-8.

Arrow, K. J., & Hurwicz, L. (1960). Decentralization and Computation in Resource Allocation. In R. W. Pfouts (Ed.), *Essays in Economics and Econometrics* (pp. 34-104). Chapel Hill: University of North Carolina Press. Reprinted in Arrow, K. J., & Hurwicz, L. (Eds.). (1977)

Arrow, K. J., & Hurwicz, L. (Eds.). (1977). *Studies in Resource Allocation Processes*: Cambridge University Press.

Barone, E. (1908). Il Ministro Della Produzione Nello Stato Collettivista. [The Ministry of Production in the Collectivist State]. *Giornale degli economisti, 37 (Anno 19)*. Translated and Reprinted in Hayek, F. A. (Ed.) (1935)

Bergson, A. (1948). Socialist Economics. In H. S. Ellis (Ed.), *A Survey of Contemporary Economics* (pp. 412-448). Philadelphia: Blakiston Co. Reprinted in Bergson, A. (1966)

Bergson, A. (1967). Market Socialism Revisited. *Journal of Political Economy, 75*, 655-673.

Blaug, M. (1997). *Economic Theory in Retrospect*. (5 ed.): Cambridge University Press.

Boettke, P. (1995). Individuals and Institutions In D. L. Prychitko (Ed.), *Individuals, Institutions, Interpretations – Hermeneutics Applied to Economics* Avebury Press.

Boettke, P. (Ed.). (2010). *Handbook on Contemporary Austrian Economics*: Edward Elgar.

Boettke, P., & Vaughn, K. I. (2002). Knight and the Austrians on Capital, and the Problem of Socialism. *History of Political Economy, 34*(1), 155-176.

Broome, J. (1990). Bolker-Jeffrey Expected Utility Theory and Axiomatic Utilitarianism. *The Review of Economic Studies, 57*(3), 477-502.

Buchanan, J. M. ([1969] 1999). *Cost and Choice: An Inquiry in Economic Theory*. Indianapolis: Liberty Fund.

Burczak, T. A. (1994). The Postmodern Moments of F. A. Hayek's Economics. *Economics and Philosophy, 10*(01), 31-58.

Caldwell, B. J. (1994). Hayek's Scientific Subjectivism. *Economics and Philosophy, 10*(02), 305-313.

Caldwell, B. J. (2005). *Hayek's Challenge: An Intellectual Biography of F.A. Hayek*. University of Chicago Press.

Chaloupek, G. K. (1990). The Austrian Debate on Economic Calculation in a Socialist Economy. *History of Political Economy, 22*(4), 659-675.

Davidson, D. (1963). Actions, Reasons and Causes. *Journal of Philosophy, 60*(23), 685-700. Reprinted in Davidson, D. (2001a)

Davidson, D. (1965). Theories of Meaning and Learnable Languages. In Y. Bar-Hillel (Ed.), *Proceedings of the International Congress for Logic, Methodology, and Philosophy of Science* (pp. 3--17): North-Holland. Reprinted in Davidson, D. (2001b)

Davidson, D. (1966). Emeroses by Other Names. *Journal of Philosophy, 63*(24), 778-780. Reprinted in Davidson, D. (2001a)

Davidson, D. (1967a). The Logical Form of Action Sentences. In N. Rescher (Ed.), *The Logic of Decision and Action*: The University of Pittsburgh Press. Reprinted in Davidson, D. (2001a)

Davidson, D. (1967b). Truth and Meaning. *Synthese, 17*(3), 304-323. Reprinted in Davidson, D. (2001b)

Davidson, D. (1969). True to the Facts. *Journal of Philosophy, 66*(21), 748-764. Reprinted in Davidson, D. (2001b)

Davidson, D. (1970). Mental Events. In L. Foster & J. W. Swanson (Eds.), *Experience and Theory*: The University of Massachussetts Press and Duckworth. Reprinted in Davidson, D. (2001a)

Davidson, D. (1971). Agency. In R. Binkley, R. Bronaugh & A. Marras (Eds.), *Agent, Action, and Reason*: University of Toronto Press. Reprinted in Davidson, D. (2001a)

Davidson, D. (1973a). Freedom to Act. In T. Honderich (Ed.), *Essays on Freedom of Action*: Routledge and Kegan Paul. Reprinted in Davidson, D. (2001a)

Davidson, D. (1973b). Radical Interpretation. *Dialectica, 27*(1), 314-328. Reprinted in Davidson, D. (2001b)

Davidson, D. (1978). Intending. In Y. Yovel (Ed.), *Philosophy of History and Action*: D. Reidel and The Magnes Press, The Hebrew University. Reprinted in Davidson, D. (2001a)

Davidson, D. (1980). A Unified Theory of Thought, Meaning, and Action. [Toward a Unified Theory of Meaning and Action]. *Grazer Philosophische Studien, 11*, 1-12. Reprinted in Davidson, D. (2004)

Davidson, D. (1987). Knowing One's Own Mind. *Proceedings and Addresses of the American Philosophical Association, 60*(3). Reprinted in Davidson, D. (2001a)

Davidson, D. (1991). Epistemology Externalized. *Dialectica, 45*(2-3), 191-202. Reprinted in Davidson, D. (2001c)

Davidson, D. (1995a). Laws and Cause. *Dialectica, 49*(2-4), 263-279. Reprinted in Davidson, D. (2005)

Davidson, D. (1995b). The Problem of Objectivity. *Tijdschrift Voor Filosofie, 57*(2), 203 - 220. Reprinted in Davidson, D. (2004)

Davidson, D. (1997). Seeing through Language. *Royal Institute of Philosophy Supplements, 42*, 15-27. Reprinted in Davidson, D. (2005)

Davidson, D. (2001a). *Essays on Actions and Events*. (2 ed.). New York: Oxford University Press.

Davidson, D. (2001b). *Inquiries into Truth and Interpretation*. (2 ed.). New York: Oxford University Press.

Davidson, D. (2001c). *Subjective, Intersubjective, Objective*. New York: Oxford University Press.
Davidson, D. (2009). *Truth and Predication*. Harvard University Press.
Davidson, D., & Suppes, P. (1956). A Finitistic Axiomatization of Subjective Probability and Utility. *Econometrica, 24*(3), 264-275.
Davidson, D., & Suppes, P. (1957). *Decision-Making: An Experimental Approach*. Stanford University Press.
Dickinson, H. D. (1933). Price Formation in a Socialist Community. *The Economic Journal, 43*(170), 237-250.
Dickinson, Z. C. (1960). Fred M. Taylor's Views on Socialism. *Economica, 27*(105), 42-52.
Dobb, M. (1939). A Note on Saving and Investment in a Socialist Economy. *The Economic Journal, 49*(196), 713-728.
Dobb, M. (1940). *Political Economy and Capitalism: Some Essays in Economic Tradition*. Routledge & Kegan Paul.
Durbin, E. F. M. (1936). Economic Calculus in a Planned Economy. *The Economic Journal, 46*(184), 676-690.
Fleetwood, S. (1995). *Hayek's Political Economy: The Socio-Economics of Order*. London: Routledge.
Geertz, C. (1974). "From the Native's Point of View": On the Nature of Anthropological Understanding. *Bulletin of the American Academy of Arts and Sciences, 28*(1), 26-45.
Gilboa, I. (2009). *Theory of Decision under Uncertainty*. Cambridge University Press.
Goodman, N. (1978). *Ways of Worldmaking*. Indianapolis: Hackett Publishing Company.
Goodman, N. (1983). *Fact, Fiction, and Forecast*. (4 ed.): Harvard University Press.
Grice, P. (1975). Logic and Conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics. 3: Speech Acts* (pp. 41-58). New York: Academic Press. Reprinted in Grice, P. (1989)
Grice, P. (1989). *Studies in the Way of Words*. Harvard University Press.
Guasti, M. T. (2004). *Language Acquisition: The Growth of Grammar*. MIT Press.
Hahn, F. H., & Negishi, T. (1962). A Theorem on Non-Tâtonnement Stability. *Econometrica, 30*(3), 463-469.
Hayek, F. A. (1935a). Socialist Calculation: The Nature and History of the Problem. In F. A. Hayek (Ed.), *Collectivist Economic Planning: Critical Studies on the Possibilities of Socialism*. London: George Routledge & Sons, Ltd. Reprinted in Hayek, F. A. (1948a)
Hayek, F. A. (1935b). Socialist Calculation: The State of the Debate. In F. A. Hayek (Ed.), *Collectivist Economic Planning: Critical Studies on the Possibilities of Socialism*. London: George Routledge & Sons, Ltd. Reprinted in Hayek, F. A. (1948a)
Hayek, F. A. (1937). Economics and Knowledge. *Economica, 4*(13), 33-54. Reprinted in Hayek, F. A. (1948a)
Hayek, F. A. (1940). Socialist Calculation: The Competitive 'Solution'. *Economica, 7*(26), 125-149. Reprinted in Hayek, F. A. (1948a)

Hayek, F. A. (1942). Scientism and the Study of Society. Part I. *Economica, 9*(35), 267-291. Reprinted in Hayek, F. A. (2010)
Hayek, F. A. (1943a). The 'Facts' of the Social Sciences. *Ethics, 54*(1), 1-13. Reprinted in Hayek, F. A. (1948a)
Hayek, F. A. (1943b). Scientism and the Study of Society. Part Ii. *Economica, 10*(37), 34-63. Reprinted in Hayek, F. A. (2010)
Hayek, F. A. (1944). Scientism and the Study of Society. Part Iii. *Economica, 11*(41), 27-39. Reprinted in Hayek, F. A. (2010)
Hayek, F. A. (1945). The Use of Knowledge in Society. *The American Economic Review, 35*(4), 519-530. Reprinted in Hayek, F. A. (1948a)
Hayek, F. A. (1948a). *Individualism and Economic Order*. The University of Chicago Press.
Hayek, F. A. (1948b). The Meaning of Competition. In F. A. Hayek (Ed.), *Individualism and Economic Order*: University of Chicago Press.
Hayek, F. A. (1952). *The Sensory Order: An Inquiry into the Foundations of Theoretical Psychology*. University of Chicago Press.
Hayek, F. A. (1963). Wirtschaft, Wissenschaft, Und Politik. [The Economy, Science And Politics]. Inaugural Lecture at the University of Freiburg, Translated and Reprinted in Hayek, F. A. (1967)
Hayek, F. A. (1967). *Studies in Philosophy, Politics, and Economics*. Chicago: University of Chicago Press.
Hayek, F. A. (2002). Competition as a Discovery Procedure. *Quarterly Journal of Austrian Economics, 5*(3), 9.
Hayek, F. A. (2010). *Studies on the Abuse and Decline of Reason: Text and Documents*. (B. Caldwell Ed.): University of Chicago Press.
Hayek, F. A. ([1941] 2007). *The Pure Theory of Capital*. Chicago: University of Chicago Press.
Hayek, F. A. ([1960] 2011). *The Constitution of Liberty*. The University of Chicago Press.
Hayek, F. A. ([1982] 2012). *Law, Legislation and Liberty: A New Statement of the Liberal Principles of Justice and Political Economy*. Routledge.
Heal, G. M. (1973). *The Theory of Economic Planning*. North-Holland Pub. Co.
Heal, G. M. (1986). Planning. In K. J. Arrow & M. D. Intriligator (Eds.), *Handbook of Mathematical Economics* (Vol. 3, pp. 1483-1510): Elsevier.
Hicks, J. (1946). *Value and Capital: An Inquiry into Some Fundamental Principles of Economic Theory*. Clarendon Press.
Houthakker, H. S. (1950). Revealed Preference and the Utility Function. *Economica, 17*(66), 159-174.
Hurwicz, L. (1969). On the Concept and Possibility of Informational Decentralization. *The American Economic Review, 59*(2), 513-524.
Jeffrey, R. C. (1983). *The Logic of Decision*. (2 ed.). Chicago: University of Chicago Press.
Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica, 47*(2), 263-291.
Keizer, W. (1987). Two Forgotten Articles by Ludwig Von Mises on the Rationality of Socialist Economic Calculation. *The review of Austrian economics, 1*(1), 109-122.

Kirzner, I. M. (1973). *Competition and Entrepreneurship*. Chicago: University of Chicago Press.
Kirzner, I. M. (1988). The Economic Calculation Debate: Lessons for Austrians. *The review of Austrian economics, 2*, 1-18.
Lange, O. (1936). On the Economic Theory of Socialism Pt. 1. *Review of Economic Studies, 4*, 53-71.
Lange, O. (1937). On the Economic Theory of Socialism Pt. 2. *Review of Economic Studies, 4*, 123-142.
Lavoie, D. (1981). A Critique of the Standard Account of the Socialist Calculation Debate. *Journal Of Libertarian Studies, 5*(1), 41-87.
Lavoie, D. (1991a). Introduction. In D. Lavoie (Ed.), *Economics and Hermeneutics*: Routledge.
Lavoie, D. ([1985] 2000). *Rivalry and Central Planning*. Routledge.
Lavoie, D. (Ed.). (1991b). *Economics and Hermeneutics*: Routledge.
Lavoie, D., & Storr, V. (2011). Distinction or Dichotomy: Rethinking the Line between Thymology and Praxeology. *The review of Austrian economics, 24*(2), 213-233.
Lawson, T. (1997). *Economics and Reality*. Routledge.
Lerner, A. P. (1934). Economic Theory and Socialist Economy. *The Review of Economic Studies, 2*(1), 51-61.
Lerner, A. P. (1936). A Note on Socialist Economics. *The Review of Economic Studies, 4*(1), 72-76.
Lerner, A. P. (1937). Statics and Dynamics in Socialist Economics. *The Economic Journal, 47*(186), 253-270.
Lerner, A. P. (1938). Theory and Practice in Socialist Economics. *The Review of Economic Studies, 6*(1), 71-75.
Lerner, A. P. (1944). *The Economics of Control: Principles of Welfare Economics*. Macmillan.
Lewis, D. ([1973] 2001). *Counterfactuals*. (2 ed.): Wiley.
Little, I. M. D. (1949). A Reformulation of the Theory of Consumer's Behaviour. *Oxford Economic Papers, 1*(1), 90-99.
Madison, G. B. (1991). Getting Beyond Objectivism: The Philosophical Hermeneutics of Gadamer and Ricoeur. In D. Lavoie (Ed.), *Economics and Hermeneutics*: Routledge.
Mäki, U. (1995). Diagnosing Mccloskey. *Journal of Economic Literature, 33*(3), 1300-1318.
Makinson, D. (2005). How to Go Nonmonotonic. In D. M. Gabbay & F. Guenthner (Eds.), *Handbook of Philosophical Logic* (Vol. 12, pp. 175-278): Springer Netherlands.
Marx, K. ([1867] 2004). *Das Kapital, Kritik Der Politischen Ökonomie*. [Capital: A Critique of Political Economy]. (B. Fowkes, Trans. E. Mandel Ed. Vol. 1): Penguin Books Limited.
Marx, K. ([1885] 2006). *Das Kapital, Kritik Der Politischen Ökonomie*. [Capital: A Critique of Political Economy]. (Vol. 2): Penguin Books Limited.

Mas-Colell, A. (1982). Revealed Preference after Samuelson. In G. Feiwel (Ed.), *Samuelson and Neoclassical Economics* (Vol. 1, pp. 72-82): Springer Netherlands.
Mas-Colell, A., Whinston, M. D., & Green, J. R. (1995). *Microeconomic Theory*. Oxford University Press.
McCloskey, D. N. (1999). *Economical Writing*. (2 ed.): Waveland Press.
Meyer, J.-J. C. (2003). Modal Epistemic and Doxastic Logic. In D. M. Gabbay & F. Guenthner (Eds.), *Handbook of Philosophical Logic* (Vol. 10, pp. 1-38): Springer Netherlands.
Mises, L. v. (1920). Die Wirtschaftsrechnung Im Sozialistischen Gemeinwesen. *Archiv für Sozialwissenschaft und Sozialpolitik, 47*, 86-121. Translated and Reprinted in Mises, L. v. ([1920] 1990)
Mises, L. v. ([1920] 1990). *Die Wirtschaftsrechnung Im Sozialistischen Gemeinwesen*. [Economic calculation in the socialist commonwealth]. (S. Alder, Trans.): Ludwig Von Mises Institute.
Mises, L. v. ([1932] 1981). *Die Gemeinwirtschaft Untersuchungen Über Den Sozialismus*. [Socialism An Economic and Sociological Analysis]. (J. Kahane, Trans. 2 ed.). Indianapolis: Liberty Fund.
Mises, L. v. ([1996] 2007). *Human Action: A Treatise on Economics*. (4 ed.). Indianapolis: Liberty Fund.
Moura, M. G. (2014). Schumpeter and the Meanings of Rationality. *FEP Working Papers*.
Munkres, J. R. (2000). *Topology*. (2 ed.): Prentice-Hall.
Neurath, O. (1919). *Durch Die Kriegswirtschaft Zur Naturalwirtschaft*. Munich: Callwey.
Neurath, O. (1932). Protokollsätze. [Protocol Sentences]. *Erkenntnis, 3*(1), 204-214. Translated and Reprinted in Ayer, A. J. (Ed.). ([1959] 1966)
Nussbaum, M. C. (2010). *From Disgust to Humanity: Sexual Orientation and Constitutional Law*. Oxford University Press.
Pareto, V. ([1909] 1966). *Manuel D'économie Politique*. (4 ed.). Genève: Droz.
Pierson, N. G. (1902). Het Waardeproblem in Een Socialistische Maatshappij. [The Problem Of Value In The Socialist Community]. *De Economist, 41*, 423-456. Translated and Reprinted in Hayek, F. A. (Ed.) (1935)
Prychitko, D. L. (Ed.). (1995). *Individuals, Institutions, Interpretations – Hermeneutics Applied to Economics* Averbury Press.
Putnam, H. (1967). The 'Innateness Hypothesis' and Explanatory Models in Linguistics. *Synthese, 17*(1), 12-22. Reprinted in Putnam, H. (1975b)
Putnam, H. (1975a). Other Minds. *Mind, Language and Reality*: Cambridge University Press.
Putnam, H. (1975b). *Philosophical Papers: Volume 2, Mind, Language and Reality*. Cambridge University Press.
Putnam, H. (1988 [1991]). *Representation and Reality*. The MIT Press.
Quine, W. V. O. (1951). Two Dogmas of Empiricism. *The Philosophical Review, 60*(1), 20-43. Reprinted in Quine, W. V. O. (1980)
Quine, W. V. O. (1964). *Word and Object*. MIT Press.

Quine, W. V. O. (1968). Ontological Relativity. *The Journal of Philosophy, 65*(7), 185-212. Reprinted in Quine, W. V. O. (1969c)
Quine, W. V. O. (1969a). Epistemology Naturalized. In W. V. O. Quine (Ed.), *Ontological Relativity and Other Essays*: Columbia University Press.
Quine, W. V. O. (1969b). Natural Kinds. In W. V. O. Quine (Ed.), *Ontological Relativity and Other Essays*: Columbia University Press.
Quine, W. V. O. (1969c). *Ontological Relativity and Other Essays*. Columbia University Press.
Quine, W. V. O. (1969d). *Set Theory and Its Logic*. (Revised ed.): Harvard University Press.
Quine, W. V. O. (1970). On the Reasons for Indeterminacy of Translation. *Journal of Philosophy, 67*(6), 178-183. Reprinted in Quine, W. V. O. (2008)
Quine, W. V. O. (1973). *The Roots of Reference*. Open Court Publishing Company.
Quine, W. V. O. (1975). On Empirically Equivalent Systems of the World. *Erkenntnis, 9*(3), 313-328. Reprinted in Quine, W. V. O. (2008)
Quine, W. V. O. (1977). Intensions Revisited. *Midwest Studies in Philosophy, 2*. Reprinted in Quine, W. V. O. (1981)
Quine, W. V. O. (1978). Goodman's Ways of Worldmaking. *The New York Review of Books*. Reprinted in Quine, W. V. O. (1981)
Quine, W. V. O. (1980). *From a Logical Point of View: 9 Logico-Philosophical Essays*. (3 ed.): Harvard University Press.
Quine, W. V. O. (1981). *Theories and Things*. Harvard University Press.
Quine, W. V. O. (2008). *Confessions of a Confirmed Extensionalist and Other Essays*. Cambridge, Massachusets: Harvard University Press.
Ramsey, F. P. (1926). Truth and Probability. In D. H. Mellor (Ed.), *Philosophical Papers*: Cambridge University Press. Reprinted in Ramsey, F. P. (1990)
Ramsey, F. P. (1990). *Philosophical Papers*. Cambridge University Press.
Rescher, N. (2005). *Epistemic Logic: A Survey of the Logic of Knowledge*. University of Pittsburgh Press.
Richter, M. K. (1966). Revealed Preference Theory. *Econometrica, 34*(3), 635-645.
Richter, M. K. (1971). Rational Choice. In J. S. Chipman, L. Hurwicz, M. K. Richter & H. F. Sonnenschein (Eds.), *Preferences, Utility, and Demand - a Minnesota Symposium*: Harcourt Brace Jovanovich, Inc.
Robbins, L. (1935). *On the Nature and Significance of Economic Science*. (2 ed.): MacMillan.
Robbins, L. ([1934] 2009). *The Great Depression*. Transaction Publishers.
Roberts, P. C. (1971). Oskar Lange's Theory of Socialist Planning. *Journal of Political Economy, 79*(3), 562-577.
Robinson, J. (1962). *Economic Philosophy*. Transaction Publishers.
Rorty, R. (1972). The World Well Lost. *Journal of Philosophy, 64*(19), 649-665.
Rorty, R. ([1979] 2009). *Philosophy and the Mirror of Nature*. (Thirtieth-Anniversary ed.): Princeton University Press.
Rothbard, M. N. (1991). The End of Socialism and the Calculation Debate Revisited. *Review of Austrian Economics, 5*(2), 51-76.
Rubinstein, A. (1998). *Modeling Bounded Rationality*. MIT Press.

Ruccio, D. F., & Amariglio, J. (2003). *Postmodern Moments in Modern Economics*. Princeton and Oxford: Princeton University Press.
Runde, J. (2001). Bringing Social Structure Back into Economics: On Critical Realism and Hayek's Scientism Essay. *The review of Austrian economics, 14*(1), 5-24.
Salerno, J. T. (1990). Postscript. In Mises, L. v. ([1920] 1990)
Salerno, J. T. (1994). Mises and Hayek on Calculation and Knowledge: Reply. *Review of Austrian Economics, 7*(2), 111-125.
Samuelson, P. A. (1938a). The Empirical Implications of Utility Analysis. *Econometrica, 6*(4), 344-356.
Samuelson, P. A. (1938b). A Note on the Pure Theory of Consumer's Behaviour. *Economica, 5*(17), 61-71.
Samuelson, P. A. (1938c). A Note on the Pure Theory of Consumer's Behaviour: An Addendum. *Economica, 5*(19), 353-354.
Samuelson, P. A. (1952). Probability, Utility, and the Independence Axiom. *Econometrica, 20*(4), 670-678.
Samuelson, P. A. ([1947] 1983). *Foundations of Economic Analysis*. (Enlarged ed.): Harvard University Press.
Savage, L. (1972). *The Foundations of Statistics*. (2 ed.): Dover Publications.
Schumpeter, J. A. (2010). *Capitalism, Socialism and Democracy*. Routledge.
Schumpeter, J. A. ([1954] 1994). *History of Economic Analysis*. Oxford University Press.
Sen, A. K. (1973). Behaviour and the Concept of Preference. *Economica, 40*, 241-259. Reprinted in Sen, A. K. (1997)
Shoenfield, J. R. (1967). *Mathematical Logic*. Association for Symbolic Logic.
Smith, A. ([1790] 1984). *The Theory of Moral Sentiments*. (6 ed.): Liberty Fund.
Smith, V. L. (2007). *Rationality in Economics: Constructivist and Ecological Forms*. Cambridge University Press.
Stalnaker, R. C. (1987). *Inquiry*. The MIT Press.
Stiglitz, J. E. (1996). *Whither Socialism?* : MIT Press.
Storr, V. (2010). The Facts of the Social Sciences Are What People Think and Believe In P. Boettke (Ed.), *Handbook on Contemporary Austrian Economics*: Edward Elgar.
Tarski, A. (1936). The Concept of Truth in Formalized Languages. In A. Tarski (Ed.), *Logic, Semantics, Metamathematics* (pp. 152--278): Oxford University Press.
Taylor, F. M. (1925). *Principles of Economics*. (9 ed.): The Ronald Press Company.
Taylor, F. M. (1929). The Guidance of Production in a Socialist State. *The American Economic Review, 19*(1), 1-8.
Tversky, A., & Kahneman, D. (1986). Rational Choice and the Framing of Decisions. *The Journal of Business, 59*(4), S251-S278.
Tversky, A., & Kahneman, D. (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty, 5*(4), 297-323.
Vaughn, K. I. (1980). Economic Calculation under Socialism: The Austrian Contribution. *Economic Inquiry, 18*(4), 535.
von Neumann, J., & Morgenstern, O. ([1944] 2007). *Theory of Games and Economic Behavior*. (Commemorative ed.): Princeton University Press.

Wakker, P. P. (2010). *Prospect Theory: For Risk and Ambiguity*. Cambridge University Press.
Wicksteed, P. H. (1910). *The Common Sense of Political Economy*. London: Macmillan and Co.
Wittgenstein, L. ([1953] 2010). *Philosophische Untersuchungen*. [Philosophical Investigations]. (G. E. M. Anscombe, P. M. S. Hacker & J. Schulte, Trans.): Wiley.
Wong, S. (1978). *Foundations of Paul Samuelson's Revealed Preference Theory, a Study by the Method of Rational Reconstruction*. Taylor & Francis.
Yeager, L. B. (1994). Mises and Hayek on Calculation and Knowledge. *Review of Austrian Economics, 7*(2), 93-109.
Yeager, L. B. (1996). Salerno on Calculation, Knowledge, and Appraisement: Rejoinder. *Review of Austrian Economics, 9*(1), 137-139.